



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



VANESSA PONTE

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DO
INTERIOR DO ESTADO DA BAHIA**

**SALVADOR
2008**



Universidade Federal da Bahia
Instituto de Letras
Programa de Pós-Graduação em Letras e Linguística

Rua Barão de Jeremoabo, nº147 - CEP: 40170-290 - Campus Universitário Ondina Salvador-BA
Tel.: (71)3283 - 6256 – Site: <http://www.ppgll.ufba.br> - E-mail: pgletba@ufba.br



VANESSA PONTE

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS POPULAR DO
INTERIOR DO ESTADO DA BAHIA**

Dissertação apresentada ao Programa de Pós-Graduação em Letras e Linguística do Instituto de Letras da Universidade Federal da Bahia como parte dos requisitos para obtenção do grau de Mestre em Letras.

Orientador: Prof. Dr. Dante Lucchesi

SALVADOR
2008

Biblioteca Central Reitor Macêdo Costa - UFBA

P813 Ponte, Vanessa.

A indeterminação do sujeito no português popular do interior do estado da Bahia / Vanessa Ponte. - 2008.
125 f.

Orientador : Prof. Dr. Dante Lucchesi.

Dissertação (mestrado) - Universidade Federal da Bahia, Instituto de Letras, 2008.

1. Língua portuguesa - Sujeito e predicado. 2. Língua portuguesa - Português falado. 3. Análise lingüística (Lingüística). I. Lucchesi, Dante. II. Universidade Federal da Bahia. Instituto de Letras. III. Título.

CDD - 469.5
CDU - 811.134.3'36

AGRADECIMENTOS

A Deus, certeza absoluta e conforto constante;

Ao professor doutor Dante Lucchesi, pela oportunidade de, desde o primeiro semestre da minha graduação, enriquecer minha vida acadêmica;

À minha mãe, pelo novo começo e pela chance de alcançar meus objetivos;

Ao meu pai, pela base da minha formação;

Aos meus tios Osvaldo e Cristina, pela acolhida em Salvador e pela ajuda no nosso recomeço;

À Vivian Antonino, amiga inseparável dentro do Projeto Vertentes, na Ufba e na vida, por toda ajuda e pela revisão do texto;

A Rerisson pela “co-orientação” e amizade todos esses anos;

À amiga Lucinda, por trilhar comigo os caminhos que me trouxeram da graduação até aqui;

À minha prima Mariucha, pela amizade sincera e constante e pela tradução do resumo;

A Luis Júnior, por todo socorro prestado;

À UFBA, pelo excelente ensino;

Às professoras da UFBA Suzana Cardoso, Therezinha Barreto, Sonia Borba, Tânia Lobo e Jacyra Andrade Mota, pelo que exigiram de mim e pelo que me ensinaram;

Às amigas do Projeto Vertentes, Luanda, Lanuza, Elisângela e Telma, sem as quais esse trabalho não poderia ser feito;

Aos meus alunos de LET A30, que contribuíram para a reflexão e aprofundamento de diversas questões acerca da constituição histórica do português do Brasil.

À CAPES, por financiar este trabalho;

RESUMO

Neste trabalho, são analisadas, segundo a metodologia da sociolinguística quantitativa, as estratégias de indeterminação do sujeito do português popular o interior do estado da Bahia. Constituem o *corpus* dessa pesquisas as entrevistas realizadas nas comunidades rurais afro-brasileiras Helvécia, Rio de Contas, Cinzento e Sapé e tanto na zona rural como na sede da cidade de Santo Antonio de Jesus. As hipóteses levantadas ao longo da dissertação são fundamentadas pelo conceito da transmissão linguística irregular e da realidade linguística bipolar, através dos quais são explicados os aspectos morfossintáticos que distinguem e, ao mesmo tempo, aproximam os dialetos rurais da norma culta. Com o estudo contrastivo dos vernáculos dessas comunidades e de comunidades rurais, portanto, pode-se mensurar de que forma o padrão culto da língua é difundido para o interior do estado. Quanto à análise dos dados, foi constatado um elenco bem maior de estratégias do que prescreve a tradição gramatical e das dez estratégias levantadas seis foram estudadas mais detalhadamente: os pronomes indeterminadores *nós*, *a gente*, *ocê* e *eles* e as formas verbais $\emptyset + V3PP$ e $\emptyset + V3PS$. Quanto aos níveis de referencialidade do sujeito indeterminado, constatou-se que há três estratégias principais para cada um desses níveis. No que concerne aos contextos linguísticos favorecedores dessas estratégias, destaca-se sobretudo a influência de dois fatores: a *forma antecedente*, única variável selecionada pelas seis estratégias de indeterminação analisadas, e o *nível de referencialidade*, em que se observou a distribuição das estratégias por seus níveis segundo a carga semântica e a capacidade de generalização. O *modo verbal* e a inclusão do falante no universo de referência do sujeito indeterminado apresentam resultados para, respectivamente, quatro e duas estratégias indeterminadoras. A influência das variáveis *tipo de frase* e *tipo de verbo* não se mostrou muito relevante. Quanto aos contextos sociais, destaca-se a influência da variável *localidade do informante* - a existência de um *continuum* linguístico é atestada pelas frequências de uso e pelos pesos relativos da forma $\emptyset + V3PS$, que apresenta maior funcionalidade nas comunidades afro-brasileiras do que nas demais localidades estudadas. Os resultados da variável *estada fora da comunidade* corroboram com a hipótese da difusão do português padrão a partir grandes centros urbanos. O caráter linguístico conservador da fala feminina é atestado pelo favorecimento da estratégia *nós*. A variável faixa etária se mostrou importante principalmente quanto às conclusões a cerca da substituição do pronome *nós* pelo *a gente*. Constatou-se também que o uso da variante padrão $\emptyset + V3PS + SE$ é extremamente reduzido na fala popular do interior do Estado da Bahia.

Palavras-chave: sujeito indeterminado; português popular; realidade linguística bipolar.

ABSTRACT

At this research, according to the quantitative sociolinguistic methodology, are analyzed the strategies of the subject indetermination in the popular Portuguese spoken at small cities of Bahia state. The *corpus* of this research are the interviews realized at the Afro-Brazilian rural communities, such as Helvécia, Rio de Contas, Cinzento, Sapé and, both in rural and in the headquarters, of the city of Santo Antonio de Jesus. The hypothesis created during the dissertation are based on the concept of the irregular linguistic transmission and the bipolar linguistic reality, through which are explained the morphosyntactical aspects that distinguish and, at the same time, approach the rural dialects to the standard portuguese. Therefore, with the contrasting study of the vernacular in these communities and rural communities we can measure how the standard language is widespread to the small cities of the state. Regarding to the data analysis, it was found a much greater cast of strategies than prescribed by the grammatical tradition and, from the ten studied strategies, six of them were studied further: the indeterminator pronouns *nós*, *a gente*, *você* and *ele* and the verbal forms $\emptyset + V3PP$ and $\emptyset + V3PS$. On the indetermined subject referentiality levels we can notice that there are three main strategies for each one of these levels. In what concern to the linguistic contexts that promotes these strategies, we identify the influence of two factors: the *antecedent history*, only variable selected by the six indetermination strategies analyzed, and the *referentiality level*, where we can observe the distribution of strategies by its levels in according to the semantic value and the capability of generalization. The *verbal mode* and the inclusion of the speaker in the universe of the indetermined subject reference, represents results to, respectively, four and two indeterminator strategies. The influence of the variables *type of phrase* and *type of verb* was not identified as relevant. At the social context, is noticed the influence of the variable *informant location* – the existence of a linguistic *continuum* is attested by the use frequency and by the form $\emptyset + V3PS$ relative influence, which presents greater functionality at the Afro-Brazilian communities than at the other studied locations. The results of the variable *being outside the community* confirm the hypothesis of the diffusion of pattern Portuguese from the big urban centers. The conservative linguistic character on the female speech is attested by the favoring of the strategy *nós*. The *age* variable was important especially on the conclusions about the substitution of the pronoun *nós*, for *a gente*. It was also observed that the use of the pattern variable $\emptyset + V3PS + SE$ is extremely reduced at popular speech from the small cities at the state of Bahia.

LISTA DE TABELAS

Tabela 1 – Formação étnica do Brasil entre 1538 e 1890	19
Tabela 2 – Dados do censo demográfico entre os anos de 1900 e 2000	26
Tabela 3 – Levantamento Geral em Milanez (1982)	33
Tabela 4 – Levantamento Geral em Milanez (1982) adaptada	34
Tabela 5 – Frequência das estratégias de indeterminação do sujeito em Rollemberg (1991)	34
Tabela 6 – Frequência dos recursos de indeterminação em Santana (2006)	38
Tabela 7 – Distribuição geral das estratégias de indeterminação do sujeito no português popular do interior do estado da Bahia	83
Tabela 8 – A indeterminação do sujeito segundo a variável realização fonética do sujeito indeterminado.	84
Tabela 9 – Estratégias de indeterminação do sujeito	85
Tabela 10 – Distribuição das estratégias de indeterminação pelos níveis de referencialidade do sujeito indeterminado	85
Tabela 11 – Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade genérico	87
Tabela 12 – Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade grupo parcialmente definido	88
Tabela 13 – Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade eu ampliado	89
Tabela 14 – Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade agente específico indefinido	90
Tabela 15 – Variáveis favorecedoras da estratégia a gente	91
Tabela 16 – O pronome a gente segundo a variável forma antecedente	91
Tabela 17 – O pronome a gente segundo a variável nível de referencialidade	92
Tabela 18 – O pronome a gente segundo a variável social localidade do informante.	92
Tabela 19 – O pronome a gente segundo a variável social faixa etária	93

Tabela 20 – O pronome a gente segundo o cruzamento das variáveis sociais localidade e faixa etária.	93
Tabela 21 – O pronome a gente segundo a variável social estada fora da comunidade.	94
Tabela 22 – Variáveis favorecedoras da estratégia nós	95
Tabela 23 – O pronome nós segundo a variável nível de referencialidade.	95
Tabela 24 – O pronome nós segundo a variável forma antecedente.	96
Tabela 25 – O pronome nós segundo a variável social localidade do informante.	96
Tabela 26 – O pronome nós segundo a variável social faixa etária	97
Tabela 27 – O pronome nós segundo a variável modo verbal.	98
Tabela 28 – O pronome nós a variável social gênero do informante.	99
Tabela 29 – O pronome nós segundo a variável social escolaridade.	100
Tabela 30 – Variáveis favorecedoras da estratégia você	101
Tabela 31 – O pronome você segundo a variável forma antecedente.	101
Tabela 32 – O pronome você segundo a variável modo verbal.	102
Tabela 33 – O pronome você segundo a variável inclusão do falante.	102
Tabela 34 – O pronome você segundo a variável social localidade do informante.	103
Tabela 35 – O pronome você segundo a variável social escolaridade.	103
Tabela 36 – Variáveis favorecedoras da estratégia eles.	104
Tabela 37 – O pronome eles segundo a variável inclusão do falante no discurso.	104
Tabela 38 – O pronome eles segundo a variável forma antecedente.	105
Tabela 39 – O pronome eles segundo a variável social localidade do informante.	105
Tabela 40 – O pronome eles segundo a variável modo verbal.	106
Tabela 41 – Variáveis favorecedoras da estratégia Ø + V3PP	106
Tabela 42 – A forma Ø + V3PP segundo a variável nível de referencialidade.	107
Tabela 43 – A forma Ø + V3PP segundo a variável forma antecedente.	107

Tabela 44 – A forma Ø + V3PP segundo a variável social localidade do informante.	108
Tabela 45 – A forma Ø + V3PP segundo a variável social estada fora da comunidade.	108
Tabela 46 – A forma Ø + V3PP segundo a variável social gênero do informante.	108
Tabela 47 – A forma Ø + V3PP segundo a variável tipo de frase.	109
Tabela 48 – A forma Ø + V3PP segundo a variável modo verbal.	109
Tabela 49 – Variáveis favorecedoras da estratégia Ø + V3PS	110
Tabela 50 – A forma Ø + V3PS segundo a variável forma antecedente.	110
Tabela 51 – A forma Ø + V3PS segundo a variável nível de referencialidade.	111
Tabela 52 – A forma Ø + V3PS segundo a variável social escolaridade.	111
Tabela 53 – A forma Ø + V3PS segundo a variável social estada fora da comunidade.	112
Tabela 54 – A forma Ø + V3PS segundo a variável tipo de verbo.	112
Tabela 55 – A forma Ø + V3PS segundo a variável social localidade do informante.	113

LISTA DE QUADROS E GRÁFICOS

Quadro 1 – Os pronomes pessoais na indeterminação do sujeito e a faixa etária.	35
Quadro 2 – Paradigmas pronominais e flexionais no português do Brasil	57
Gráfico 1 – Os pronomes nós e a gente e a variável faixa etária	97
Gráfico 2 – O pronome nós e o cruzamento das variáveis gênero e localidade do informante.	99

SUMÁRIO

LISTA DE TABELAS

LISTA DE QUADROS E GRÁFICOS

INTRODUÇÃO	12
1 O PORTUGUÊS POPULAR DO BRASIL	16
1.1 POVOS EM CONTATO	16
1.1.1 Os indígenas	17
1.1.2 Os africanos	18
1.2 A POLÊMICA EM TORNO DA FORMAÇÃO DO PORTUGUÊS POPULAR	20
1.3 A REALIDADE LINGÜÍSTICA BIPOLAR	24
2 A INDETERMINAÇÃO DO SUJEITO	28
2.1 O ENFOQUE DA GRAMÁTICA NORMATIVA	29
2.2 ANÁLISE DO TEMA NO ÂMBITO DA LINGÜÍSTICA MODERNA	31
2.3 OS GRAUS DE INDETERMINAÇÃO DO SUJEITO	38
2.4 DETERMINAÇÃO, INDETERMINAÇÃO E INDEFINIÇÃO	43
2.4.1 Indeterminação e determinação	43
2.4.2 Indeterminação e indefinição	44
2.5 A PARTÍCULA SE	44
3 FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA	47
3.1 A TEORIA	47
3.1.1 A Pesquisa sociolingüística	47
3.1.2 A Transmissão lingüística irregular e o português do Brasil	49
3.1.3 O sujeito pronominal	51

3.1.3.1 O parâmetro do sujeito nulo	51
3.1.3.2 O sujeito pronominal e as línguas crioulas	54
3.1.3.3 O comportamento sujeito pronominal no português do Brasil	56
3.1.4 Hipóteses acerca da indeterminação do sujeito no português popular	58
3.2 METODOLOGIA DE TRABALHO	59
3.2.1 As comunidades estudadas	60
3.2.1.1 Breve histórico das comunidades afro-brasileiras	60
3.2.1.1.1 A comunidade de Helvécia	60
3.2.1.1.2 A comunidade de Rio de Contas	62
3.2.1.1.3 A comunidade de Sapé	63
3.2.1.1.4 A comunidade de Cinzento	65
3.2.1.2 Breve histórico da cidade de Santo Antonio de Jesus	65
3.2.2 O corpus	67
3.2.2.1 As entrevistas	67
3.2.2.2 Os informantes	68
3.2.3 O suporte quantitativo	68
3.2.4 Levantamento das ocorrências	69
3.2.5 Variáveis lingüísticas	71
3.2.5.1 Realização fonética do sujeito indeterminado	72
3.2.5.2 Desinência verbal	72
3.2.5.3 Tipo de Verbo	73
3.2.5.4 Tipo de frase	74
3.2.5.5 Nível de referencialidade	75
3.2.5.6 Modo verbal	76
3.2.5.7 Inclusão do falante	76
3.2.5.8 Forma antecedente	77
3.2.6 Variáveis sociais	79
3.2.6.1 Localidade do falante	79
3.2.6.2 Gênero	79
3.2.6.3 Faixa etária	81

3.2.6.4 Estada fora da comunidade	81
3.2.6.5 Escolaridade	82
4 A ANÁLISE DOS RESULTADOS	83
4.1 OS NÍVEIS DE REFERENCIALIDADE DO SUJEITO INDETERMINADO	85
4.1.1 Nível genérico	86
4.1.2 Grupo parcialmente definido	87
4.1.3 Eu ampliado	88
4.1.4 Agente específico indefinido	89
4.2 A INDETERMINAÇÃO DO SUJEITO NO INTERIOR DO ESTADO DA BAHIA	90
4.2.1 A gente	90
4.2.2 Nós	95
4.2.3 Você	100
4.2.4 Eles	104
4.2.5 Ø + 3VPP	106
4.2.6 Ø + 3VPS	110
4.3 O SE	114
5 CONSIDERAÇÕES FINAIS	117
REFERÊNCIAS	121

INTRODUÇÃO

Para as Gramáticas Tradicionais (GT), a língua portuguesa dispõe de duas estratégias para indeterminar o sujeito: construções com o sujeito nulo e verbo em terceira pessoa do plural, como exemplificado em (01), e verbos intransitivos, transitivos indiretos ou transitivos tomados como intransitivos em terceira pessoa do singular acompanhados da partícula *se*, representados, respectivamente, pelos exemplos (02), (03) e (04).¹

(01) **Atropelaram** uma senhora na esquina.

(02) Aqui **vive-se** bem.

(03) **Precisa-se** do carvalho; não se precisa do caniço.

(04) **Comia-se** com a boca, com os olhos, com o nariz.

Poucos são os autores que divergem desse pensamento, como Bechara (1977) e Cegalla (1997), por exemplo, que acrescentam às duas estratégias já citadas uma terceira. Entretanto os autores não compartilham do mesmo posicionamento: Bechara cita o uso do verbo na terceira pessoa do singular, no exemplo (05), enquanto Cegalla cita o infinitivo pessoal, no exemplo (06).

(05) **Diz** que eles vão bem.

(06) É triste **assistir** a estas cenas repulsivas.

Quanto ao uso da partícula *se*, as GTs consideram erro a ausência de concordância entre o sujeito paciente e o verbo, em estruturas como a representada em (07).

(07) **Aluga-se** apartamentos.

¹ Exemplos de Cegalla (1997) e Cunha e Cintra (2001).

Pesquisas que têm a indeterminação do sujeito, na língua oral e na escrita, como foco (Cf. MILANEZ, 1982; CUNHA, 1993; CAVALCANTE, 1999; SANTANA, 2006) e mesmo a análise mais aprofundada do comportamento do sujeito pronominal (Cf. DUARTE, 1995; ILARI *et al*, 1996; NEVES, 2000) identificaram, além das já citadas, mais estratégias de indeterminação do sujeito e revelaram serem as estratégias pronominais, representadas em (08), (09) e o sujeito nulo, em (10), abaixo, mais produtivas que as estratégias padrão.

(08) Porque eu acho assim, se *a gente*... é... maltratar uma pessoa dessa, *a gente* só vai esperar o que deles?

(09) Porque quanto mais *you* estuda, mais incentivo *you* 'tá dando pro seus filho.

(10) Um dinheirinho já serve. Ø Já sabe o que fazer com aquele dinheiro.

Compartilhando esse objetivo de descrever as reais estratégias de indeterminação do sujeito no português, esta dissertação se baseia na análise sociolinguística do comportamento do sujeito indeterminado em variedades distintas do português popular brasileiro, como as que se encontram em comunidades afro-brasileiras isoladas e num município do interior, localizadas no estado da Bahia.

Desta forma, esta pesquisa teve por objetivos:

- a) descrever as estratégias de indeterminação do sujeito no português popular do interior do estado da Bahia;
- b) verificar os fatores linguísticos e sociais que condicionam a escolha pelas variantes;
- c) contrastar o comportamento linguístico, no que concerne à indeterminação, no vernáculo de comunidades afro-brasileiras isoladas e no português popular de um município do interior do estado.

Para tanto, foram levantadas e analisadas quantitativamente as ocorrências em orações finitas dos pronomes *a gente*, *nós*, *you* e *eles* sem uma referência específica e das formas verbais da terceira pessoa do singular e do plural com o pronome sujeito não realizado foneticamente e sem ligação com uma referência anterior específica.

A hipótese central do trabalho é a de que o processo sócio-histórico pelo qual passou o português trazido para o Brasil teria deixado marcas da aprendizagem defectiva por parte

dos africanos e seus descendentes, principais difusores da língua portuguesa no país (MATTOS E SILVA, 2004), e de que estas marcas seriam presentes, sobretudo, nos falares das comunidades afro-descendentes isoladas, como as estudadas pelo *Projeto Vertentes*, no âmbito do qual este trabalho foi desenvolvido². O contraste entre o vernáculo das comunidades afro-brasileiras e o português popular de um município do interior do estado confirmaria ou não a hipótese de um *continuum* lingüístico proveniente dos processos sócio-históricos e lingüísticos englobados pelo conceito de *transmissão lingüística irregular* (LUCCHESI, 2003a), contribuindo para o esclarecimento da realidade lingüística do país. Assim, o reconhecimento da realidade lingüística brasileira como *bipolarizada* que fundamenta esta análise (LUCCHESI, 2001) é crucial para o seu desenvolvimento argumentativo, que se estrutura da seguinte maneira.

No capítulo um, são abordados os processos sócio-históricos responsáveis pela relevância do elemento africano na difusão do português no Brasil, bem como o posicionamento de diversos estudiosos acerca, primeiramente, da influência dos falares das etnias não-brancas e, em seguida, das origens do português popular do Brasil (doravante PPB). É tratada também a hipótese da bipolarização da realidade lingüística do português do Brasil (PB).

No segundo capítulo, são analisadas as descrições da indeterminação do sujeito em algumas Gramáticas Tradicionais; também são apresentados os resultados obtidos por diversos pesquisadores. *Os níveis de referencialidade* do sujeito indeterminado, variável tratada em outros trabalhos como graus de indeterminação do sujeito, tem destaque nesta sessão.

O terceiro capítulo, intitulado *Fundamentação teórica e metodologia*, é dividido em duas partes. A primeira apresenta os pressupostos teóricos da *Sociolingüística Variacionista Quantitativa*, ou Laboviana, que fundamentam este trabalho; são feitas, ainda na primeira parte, considerações sobre o comportamento da indeterminação do sujeito no PB e sua relação com o processo de transmissão lingüística irregular. O *Método* é apresentado na segunda parte do capítulo, em que se descreve o caminho traçado para o desenvolvimento da pesquisa no âmbito da Sociolingüística Quantitativa.

O capítulo quatro apresenta os resultados quantitativos obtidos ao longo desta pesquisa e tece as considerações necessárias acerca dos fatores que condicionam o uso de cada uma das variantes analisadas.

² Para maiores informações sobre o Projeto *Vertentes do Português Rural do Estado da Bahia*, veja-se a página do Projeto na Internet: <http://www.vertentes.ufba.br/index.htm>

Na sessão *Considerações Finais*, é apresentado um resumo dos resultados obtidos e são ratificados os pontos mais importantes do trabalho. São também apresentados futuros desdobramentos para esta pesquisa.

1 O PORTUGUÊS POPULAR DO BRASIL

As línguas são produtos históricos e, portanto, para o estudo de uma língua torna-se necessário um aprofundamento não só na história interna da língua, mas também a análise da sua história externa.

Os processos que balizaram a formação do PB, bem como a caracterização da realidade lingüística do país são alvos de inúmeros trabalhos desde o início do século XX e estão no centro da polêmica sobre qual teria sido a influência do contato entre línguas na formação das variedades populares do português. De um lado, lingüistas como Serafim da Silva Neto (1963), Mattoso Câmara Jr. (1975), Silvio Elia (1979), Anthony Naro e Marta Scherre (1993a) se posicionam a favor da deriva secular da língua, uma tendência natural e prefigurada que teria sido somente apressada em função desse contato entre povos; do outro lado, estão Jacques Raimundo (1933), Renato Mendonça (1973 [1933]), Gregory Guy (1981, 1989) e Dante Lucchesi (1994, 2001 e 2003a), pesquisadores que defendem a relevância do contato entre línguas na formação das variedades populares do português.

Neste capítulo, além de uma breve descrição dos processos sociais e históricos pelos quais passou o português e das personagens que contribuíram para a configuração atual desta língua, serão apresentados os posicionamentos de alguns estudiosos acerca da formação do PPB. Finalmente, será debatida a hipótese da bipolarização da realidade lingüística brasileira, que fundamenta esta análise.

1.1 POVOS EM CONTATO

Nesta sessão, serão analisados, mesmo que brevemente, aspectos sociais e históricos que concernem a dois grupos que contribuíram para a implementação e difusão da língua portuguesa em território brasileiro: os indígenas e os negros africanos. A ação predatória da violência cultural e do extermínio de que foram alvo os indígenas, dentre outros fatores que também serão tratados, fizeram com que os africanos e afro-descendentes se tornassem os principais difusores do PPB³; e é esse assunto que será abordado a seguir.

³ Toma-se o termo *popular*, neste trabalho, com a acepção de *não-padrão*, em oposição ao português culto ou padrão.

1.1.1 Os indígenas

Na ocasião da chegada dos colonizadores portugueses às terras que futuramente formariam o Brasil, as línguas indígenas faladas pelas tribos que aqui habitavam eram, segundo Ayron Rodrigues (1993, p. 93), em torno de 1.175, das quais 85% desapareceram somente no período colonial. Ao longo da costa, as línguas faladas procediam do tronco tupi e eram muito aparentadas, a comunicação era facilitada, pois esses falantes utilizavam uma *língua geral*, termo definido por Rodrigues (1996) como “um termo específico para determinada categoria de línguas, que surgiram na América do Sul nos séculos XVI e XVII em condições especiais de contato entre europeus e povos indígenas”.⁴ Conforme o autor, as línguas gerais distinguem-se de línguas *pidgins* e dos *crioulos* pelo contexto sociolinguístico que propiciaram o seu surgimento, marcado pelo bilingüismo parcial e pelo convívio de casais mistos (Cf. RODRIGUES, 1996, p. 8).

O contato entre homens europeus e as mulheres indígenas em grande escala (favorecido pela prática do cunhadismo, comum entre os indígenas) resultou em populações predominantemente mestiças e propiciou o surgimento no Brasil de duas línguas gerais: a língua geral paulista, surgida no século XVI entre os portugueses e os índios tupiniquins em São Vicente e no planalto de Piratininga, no que seria o leste do atual estado de São Paulo; e a língua geral amazônica, produto do contato entre portugueses e tupinambás no século XVII, no norte dos atuais estados do Maranhão e do Pará.

A língua geral toma conta das terras brasileiras e a larga utilização desta língua leva os jesuítas a utilizá-la na catequese do gentio.

O processo colonizador e evangelizador dos séculos XVI e XVII teve de utilizar, como instrumento fundamental para a dominação, línguas indígenas brasileiras. Embora homogeneizadora da atividade catequética – construiu um “tupi jesuítico” – a língua geral da costa, de base tupi, chegou a ser um risco para a hegemonia do português no Brasil, juntamente com outras línguas gerais indígenas que foram veículos de intercomunicação entre brancos, negros e índios não só no litoral brasileiro [...]. (MATTOS E SILVA, 2004, p. 14-15)

A força de trabalho indígena foi utilizada primeiramente na extração do pau-brasil e, mais tarde, no cultivo da cana-de-açúcar, do tabaco e do algodão. O indígena, porém, devido a sua organização social, na qual homens, mulheres e crianças possuem tarefas distintas, a-

⁴ A expressão *língua geral* foi utilizada tanto em referência às línguas do tronco tupi-guarani, quanto para referir-se à língua moldada pelos jesuítas no século XVI. Segundo Lucchesi (2000, p. 43-44), o termo recobre cinco situações sociolinguísticas distintas.

presentou uma profunda resistência ao trabalho forçado. A política de colonização empregada pelos colonizadores, que resultava no progressivo extermínio e na escravização dos indígenas nos chamados ciclos de apresamento, propiciou também, juntamente com a extinção de diversos povos, a extinção de inúmeras línguas indígenas, das quais nunca se terá o conhecimento.

Apesar da extensão territorial alcançada pela língua geral paulista em função da expansão paulista favorecida pelas bandeiras e pela busca por escravos, pouca documentação há sobre esta língua. Na segunda metade do século XVIII, medidas legislativas e administrativas foram impostas pelo Marquês de Pombal para estimular o uso do português e combater o uso da língua geral. Estas medidas surtiram efeito sobre a língua geral paulista e contribuíram para o enfraquecimento do uso das línguas indígenas.

As campanhas jesuíticas contra a escravidão do gentio, a dificuldade encontrada pelos colonizadores em trabalhar com a resistência cultural do índio ao trabalho forçado e as dificuldades na captura de novas “peças”, que para fugir adentravam cada vez mais o território ainda pouco explorado, tornou necessário que se buscasse outras fontes de mão-de-obra.

1.1.2 Os africanos

Em 1549, com a autorização de Dom João III, o tráfico negreiro é oficializado no Brasil ao limite de 120 “peças” para cada engenho montado, contudo, antes da oficialização, já eram “solicitados escravos negros pelo donatário Duarte Coelho da capitania de Pernambuco, em 1539, e pelo donatário da capitania de São Tomé, Pero de Góis, em 1545” (MATTOS E SILVA, 2004, p. 96).

Com os africanos trazidos ao Brasil, chegam cerca de 200/300 línguas africanas, segundo estimativas de Emílio Bonvini e Margarida Petter (1988 Apud MATTOS E SILVA, 2004, p. 96). Estas línguas são provenientes de duas grandes regiões subsaarianas: a área banto, que abrange toda a extensão abaixo da linha do Equador, e a área da África ocidental, situada ao norte da linha do equador. No primeiro grupo, predominam as línguas étnicas *quimbundo*, *quicongo* e *umbundo*; no segundo, os principais representantes são os *nagôs* ou *iorubás* e os *jejes*, ou povos de língua *ewe*.

A situação de multilingüísmo em que o Brasil se encontrava até então foi intensificada em função das inúmeras línguas africanas. A língua geral ainda era falada, porém perdia espaço nos grandes centros urbanos e se restringia ao interior do país.

Assim, enquanto em São Paulo, no Maranhão e no interior do país (ou seja, nas zonas periféricas da colônia), a língua geral predominava em função da submissão e aculturação das populações indígenas, a língua portuguesa avançava a partir da Bahia e Pernambuco, os centros mais dinâmicos e mais intimamente ligados à economia mercantilista, na qual se integrava o projeto colonial brasileiro. (LUCCHESI, 2000, p. 45)

Os escravos não eram agrupados pela língua, no intuito de evitar rebeliões. Era necessário, portanto, que aprendessem uma outra língua que pudesse ser utilizada para a comunicação entre tantos povos falantes de línguas diversas. O aprendizado, é claro, era feito na oralidade e sem normatização.

A falta de registros oficiais, entretanto, dificulta determinar tanto a procedência quanto a quantidade dos africanos para cá trazidos. Estima-se que tenha sido em torno de 3.500.000 (MATTOSO, 1990, p. 13), mas Darcy Ribeiro chega ao número de 6.353.000 escravos chegados entre 1540 e 1860.

Quanto à composição demográfica da população do Brasil, Tânia Lobo (1996 Apud MATTOS E SILVA, 2004, p. 102), ao reorganizar os dados de Alberto Mussa (cf. tabela 1), deixa clara a importância dos africanos na composição étnica do país e a predominância de línguas outras que não a portuguesa. Com exceção do último período de tempo disposto na tabela, a etnia branca não ultrapassou um terço da população total do país.

Tabela 1 – Formação étnica do Brasil entre 1538 e 1890

	etnias não-brancas	etnia branca
1538-1600	70%	30%
1601-1700	70%	30%
1701-1800	68%	32%
1801-1850	69%	31%
1851-1890	59%	41%

Fonte: Lobo (1996, p. 16)

Diante desses dados demográficos e da progressiva e crescente extermínio sofrida pelos povos indígenas, admite-se não ter sido o indígena, mas sim o africano e o afrodescendente um difusor do português geral brasileiro, antecedente histórico do atual português popular (MATTOS E SILVA, 2004, p. 102), no Brasil.

Outros tantos fatores como mobilidade geográfica dos escravos e a sua presença nas grandes frentes de exploração econômica também contribuem para este posicionamento. No século XVI e até início do século XVII, as lavouras canavieiras de Pernambuco, Bahia e Rio

de Janeiro eram responsáveis pela expansão da língua portuguesa no Brasil, através “da fala dos colonos portugueses que afluíam para esse empreendimento e, sobretudo, através das variedades defectivas de português adquiridas pelos escravos africanos e transmitidas para os seus descendentes crioulos” (LUCCHESI, 2000, p. 48). No final do século XVII e no decorrer do século XVIII, a descoberta de jazidas de ouro e pedras preciosas em Minas Gerais, Mato Grosso e Goiás dá início ao ciclo do ouro e redireciona a mão de obra escrava para as frentes de mineração. Os senhores de engenho decadentes vendiam seus escravos para os exploradores de jazidas ou para lá se direcionavam buscando obter riquezas. A partir de 1763, a cidade do Rio de Janeiro passa a ser a capital da colônia e o crescimento da vida urbana torna a cidade o maior porto importador de escravos do país. Há um novo impulso açucareiro na área que vai do Rio de Janeiro a São Paulo e a mão de obra é atraída para o litoral. A economia do país continua sendo fundamentalmente agro-exportadora e a mão de obra, em sua grande maioria, escrava. No século XIX, o foco da economia do país eram as fazendas cafeeiras do Vale do Paraíba, que abrangiam as regiões de São Paulo, Rio de Janeiro e Minas Gerais.

Os papéis desempenhados pelos escravos na sociedade brasileira não se restringiam aos trabalhos forçados ou à senzala. Na casa grande, nas grandes propriedades rurais, podiam ser encontrados escravos domésticos ou amas, que participavam diretamente da criação dos filhos dos seus senhores, influenciando-lhes o falar com as marcas da aquisição defectiva do português. Os quilombos também eram indícios significativos do dinamismo social dos negros e afro-descendentes; neles era possível encontrar não só ex-escravos, mas também índios e fugitivos de qualquer etnia. A configuração lingüística desses espaços ilegítimos da escravidão era diversa a depender da sua localização e composição.

Todos esses fatores sociais e históricos são decisivos na defesa da hipótese de que seriam o elemento negro e os seus descendentes os principais difusores da língua portuguesa no Brasil.

1.2 A POLÊMICA EM TORNO DA FORMAÇÃO DO PORTUGUÊS POPULAR

A origem do português popular está no centro de um debate que se iniciou na segunda metade do século XIX, motivado pela independência do país e pelo movimento romântico, e até hoje perdura. A geração romântica inicia, dentro da literatura, a constituição de uma língua nacional, que simbolizaria uma ruptura com a tradição literária portuguesa. Porém seriam

os modernistas que buscariam na realidade lingüística brasileira a independência da língua do Brasil com relação à língua da ex-metrópole, retomando a defesa da existência de uma língua brasileira.

Ainda na primeira metade do século XX, a questão da língua portuguesa no Brasil migra do meio literário para o lingüístico, e outra questão, além da autonomia da língua do Brasil perante a língua de Portugal, torna-se tema de uma polêmica: a influência das línguas africanas sobre o português popular. Os posicionamentos de diversos autores divergem: enquanto alguns estudiosos defendem que os falares africanos nada mais fizeram do que acelerar a ação das forças internas já prefiguradas na língua, outros buscam as justificativas para a configuração do PPB na constituição sócio-histórica do país.

Em 1933, Jacques Raimundo, no livro *O elemento afro-negro no Brasil*, observa algumas alterações na pronúncia do PB em processos mórficos e no arranjo das frases e atribui esses fenômenos à influência das línguas africanas. O estudo, ainda, é centrado no estudo do léxico, como uma primeira tentativa para classificar e analisar a contribuição dos africanos. São apresentadas em seu trabalho 309 palavras, além de 132 topônimos.

Ainda no mesmo ano, Renato Mendonça atribui as características que distinguem o português europeu do brasileiro à influência das línguas africanas. Dentre as alterações sofridas pela língua portuguesa, o autor cita fenômenos relacionados à morfologia, como a ausência de plural no substantivo e no adjetivo na função predicativa, quando a marca vem expressa no determinante. O maior destaque, porém, é dado às alterações fonéticas. Já, na sintaxe, essa influência seria ainda menor do que a verificada na morfologia. Mendonça (1973 [1933]) procura também comparar as ocorrências desses fenômenos com as ocorrências em crioulos e línguas românicas.

Gladstone Chaves de Melo publica em 1946 *A língua portuguesa do Brasil* e afirma ter sido a influência dos africanos do PPB maior do que a dos índios, já que viveram um longo contato com o homem branco. As alterações sofridas pelo português, entretanto, são classificadas como “deturpações”. Melo afirma ter existido no Brasil um crioulo de base tupi-quechua que teria “deformado” o português arcaico. Além disso, atribui a esse último diversas características encontradas no PB, já atribuídas por Jacques Raimundo e Renato Mendonça à influência dos falares africanos. Contudo não encontra no português arcaico correspondente para a redução na morfologia flexional.

Em 1950, Serafim da Silva Neto, em sua *Introdução ao estudo da língua portuguesa no Brasil* afirma serem as principais características do PB a unidade e o conservadorismo. O

autor atribui aos africanos somente a realização de algo que já estaria pré-configurado na língua portuguesa (SILVA NETO, 1988 [1957]), crendo numa aceleração da deriva secular da língua. No caso do PB, Silva Neto diz ter havido duas derivas, uma conservadora e uma inovadora. Os falantes portugueses seriam responsáveis pelos traços conservadores, enquanto os africanos, pela aquisição imperfeita da língua, seriam responsáveis pela inovação. A justificativa do autor para a presença de ambos os traços tanto na variedade culta como na popular está na “passividade receptiva” do crioulo, que coincidiu com o purismo da classe média. Quanto ao posicionamento sócio-histórico e ideológico do autor, toda a obra é permeada pela convicção da superioridade étnica e cultural do colonizador, o que pode ser percebido claramente em trechos como “[...] no decurso de quatro séculos, umas linguagens reagiram sobre outras, sempre no sentido da preponderância daquela que representava o mais alto e perfeito meio de civilização: o português (SILVA NETO, 1963, p. 67)”.

Tânia Lobo (1994) analisa três aspectos da tese defendida por Silva Neto com o intuito de verificar a pertinência da sua argumentação. Segundo Lobo, o estágio da lingüística no Brasil, as análises acerca da história interna da língua, mantidas dentro da fronteira da pronúncia, e o condicionamento ideológico do autor, refletido no seu posicionamento diante dos fatos históricos, não seriam suficientes para a defesa da tese da unidade e do conservadorismo do PB.

A defesa da tese da unidade e do conservadorismo do português brasileiro, ou até mesmo da tese contrária, só se poderia legitimar se precipuamente baseada em uma sistemática verificação empírica dos dados lingüísticos, o que, não tendo ocorrido, lhe confere um caráter impressionístico (LOBO, 1994, p. 12).

Assim como Silva Neto, Matoso Câmara Júnior (1975) também nega um papel relevante aos falares africanos, admitindo ter havido somente a aceleração da deriva secular da língua portuguesa. Para o autor, a existência de arcaísmos na variedade brasileira do português, característica marcante, se justificaria pelo isolamento de determinadas áreas com relação aos centros urbanos da colônia.

Silvio Elia (1979) defende a unidade e o conservadorismo do PB. Apesar de admitir a influência “afro-índia”, diz que o falar constituído no Brasil é uma “deturpação ou simplificação de uma língua-base, quando falada por aloglotas de estágio cultural inferior” (p. 234). O autor afirma que o contato entre portugueses, africanos e índios não alterou o sistema do português. Contudo são admitidas influências quanto à simplificação da morfologia, à alteração da prosódia e à sintaxe.

Em 1981, Gregory Guy assume a posição de que o PB teve origem crioula. Em seu estudo, a variedade brasileira do português é comparada com línguas crioulas de base lexical portuguesa e, a partir disso, o autor verifica ser esta uma característica comum entre essas línguas. Os dialetos rurais apresentariam fortes indícios da crioulição sofrida pelo PB, em vias de descrioulização, propiciada pelo processo de urbanização do país. Fatores sociais, tais como a faixa etária, seriam desencadeadores da aquisição das marcas de concordância.

No artigo *Sobre a legada origem crioula do português brasileiro*, Fernando Tarallo (1993) defende que na discussão da possível origem crioula do PB são cruciais tanto a análise da história interna da língua quanto da história externa, mas só a última fornece dados que comprovariam esta hipótese. O autor concentra-se na idéia de que o PB não estaria passando por um processo de descrioulização e empenha-se para provar que as mudanças sintáticas a serem discutidas são aleatórias e não caminham em direção ao PE. A análise de dois fatores constitui a base da sua argumentação: a retenção dos pronomes em sentenças matrizes e encaixadas, alta no PE e baixa no PB, e a retenção dos pronomes em posição de objeto, que se realiza frequentemente no PB, mas não no PE; e levam-no a concluir que estas duas variedades lingüísticas seguem em direções opostas. Conforme Fernando Tarallo, para que o PB se descrioulizasse em direção ao PE, seria necessário que virasse “pelo avesso e de ponta cabeça” (1993, p. 60).

Também no ano de 1993, Anthony Naro e Marta Scherre discutem a origem do PPB fazendo uso da deriva secular indo-européia para justificar as simplificações das flexões no PB. Desta forma, compartilham da posição de Silva Neto de que as distinções entre PB e PE são reflexos da aceleração de uma tendência já pré-configurada na língua. Os autores não negam que tenha havido transmissão lingüística irregular, mas assumem a postura de que esta não foi o que propiciou as alterações na concordância verbal e nominal apresentadas pelo PB. Quanto à questão da influência das línguas africanas em Naro e Scherre (1993), Silva (2005) observa:

A afirmação de Naro e Scherre de que “os observadores – nativos ou estrangeiros – não conseguem reconhecer qualquer traço lingüístico associado exclusivamente com a etnia afro-brasileira” (NARO; SCHERRE, 1993, p. 438-9) não pode ser aceita se considerarmos o silêncio acadêmico que até o século XX pairou sobre a figura dos escravos africanos e seus descendentes (p. 31-32).

Baxter e Lucchesi (1997) defendem a importância de se estudar o contato entre línguas no Brasil e destacam aspectos gramaticais do dialeto da comunidade afro-descendente de Helvécia, no extremo sul da Bahia, provenientes da aprendizagem imperfeita do português

pelos africanos e seus descendentes. O conceito de língua crioula é revisitado neste trabalho e suas características sociolingüísticas, assim como suas particularidades estruturais, são apresentadas para que, a partir de uma definição atualizada, possam ser tecidos comentários acerca da relevância dos processos de criouliização na formação do PB. Desta forma, segundo os autores

[...] é melhor contemplar a criouliização como um processo que pertence a um contínuo de gêneros de transmissão de L2 para L1, partindo do mais irregular para o mais regular, a depender da qualidade da L2 que serviu de base para o surgimento da língua crioula, e, portanto, das circunstâncias sócio-históricas da transmissão (p. 74).

Os aspectos morfossintáticos dos dialetos rurais, então, são explicados em termos dessa reflexão. Assim, a variação na concordância de número no sintagma nominal e a redução da morfologia verbal de pessoa e número, por exemplo, presentes no dialeto rural de Helvécia, são tomados como características crioulizantes.

Seguindo a mesma linha, Lucchesi (2003a) formaliza o conceito de *transmissão lingüística irregular* através da definição de parâmetros sócio-históricos e lingüísticos que abrangeriam as situações de contato entre línguas que não resultam na formação de pidgins e crioulos típicos. O conceito de transmissão lingüística irregular designa “processos históricos de contato massivo entre línguas nos quais a língua do segmento que detém o poder político é tomada como modelo ou referência para os demais segmentos” e “constitui um contínuo de níveis diferenciados de socialização/nativização de uma língua segunda” (Lucchesi, 2003a, p. 272). Assumindo a postura de que “em todo e qualquer processo de transmissão lingüística irregular desencadeado pelo contato entre línguas ocorre, em maior ou menor grau, perda de morfologia flexional e de regra de concordância nominal e verbal” (p. 282), Lucchesi afirma ter o contato entre língua um papel relevante na história da língua portuguesa no Brasil.

1.3 A REALIDADE LINGÜÍSTICA BIPOLAR

A tese da bipolarização da realidade lingüística brasileira (Cf. Lucchesi, 1994, 1996, 2001) encontra sua fundamentação no processo sócio-histórico de constituição da sociedade brasileira. Nos primeiros séculos da história do Brasil, destacavam-se dois pólos de comportamentos lingüísticos distintos: de um lado, uma pequena elite colonial de comportamento lingüístico conservador; e do outro lado, o pólo das camadas populares, em que o português

era modificado em decorrência do contato do português com as línguas indígenas e africanas. Enquanto a elite colonial se voltava para os padrões lingüísticos e culturais da metrópole, o português era aprendido precariamente e disseminado nas camadas populares.

Durante o período colonial, a maior parte da população se dirigiu para o interior do país. A elite se restringia a pequenos centros urbanos litorâneos. No interior, a língua portuguesa convivía com variedades da língua geral, língua franca de base tupinambá, e, posteriormente, com predominância da mão-de-obra escrava no Brasil, com as línguas africanas. A mobilidade geográfica dos africanos e afro-descendentes, além da sua grande densidade demográfica, contribuiu para fundamentar a hipótese de que teriam sido esses os principais difusores da língua portuguesa no Brasil (MATTOS E SILVA, 2004). Após o declínio da cultura da cana-de-açúcar no nordeste, o sudeste passa a ser o centro econômico do país, devido à descoberta de grandes jazidas de ouro e pedras preciosas; juntamente com os escravos, um grande número de mestiços e pequenos comerciantes, criadores de gado e lavradores se deslocam para o interior do Brasil. Com o esgotamento dessas jazidas, as plantações de café do Vale do Rio Paraíba absorvem a mão-de-obra escrava, que novamente deslocava-se.

Aprendida e difundida pela população pobre de origem predominantemente indígena e africana, a língua portuguesa sofria drásticas transformações que acentuavam o abismo existente entre as variedades populares e a variedade cultivada pela aristocracia.

Contudo a hegemonia da língua portuguesa se estabelece gradativamente ao longo da colonização do país, e só a partir de meados do século XIX a situação de multilingüismo que predominava até então dá lugar ao unilingüismo (LOBO, 2003, p. 403).

No fim do século XIX, quando se intensifica o processo de industrialização e urbanização do país, o panorama sócio-histórico e demográfico é profundamente modificado. A maciça migração rural aponta para um recente e intenso processo de urbanização do país. A tabela 2 apresenta dados do censo demográfico recolhidos por Bortoni-Ricardo (2005, p. 92) que demonstram esse processo recente, sensível a partir dos anos 40 e 50, mas implementado em todo país na década de 80.

Tabela 2 – Dados do censo demográfico entre os anos de 1900 e 2000.

Ano	População total	População urbana (%)
1900	14.333.915	6,8
1920	-	10,7
1940	-	31,29
1950	51.944.397	36,16
1980	120.000.000	67,60
1991	157.000.000	78,35
2000	168.370.893	81,37

Fonte: Bortoni-Ricardo (2005)

No fim do século XIX, a imigração, preferencialmente de europeus, era estimulada pelo rápido desenvolvimento dos Estados Unidos, que já recebiam imigrantes há décadas, pela crença na superioridade racial das etnias brancas e pelo momento histórico de ênfase da formação da nacionalidade. Apesar da aparente contradição entre afirmação da nacionalidade e incentivo da imigração, o que era buscado era “um pretense coletivo, operava-se uma universalização do conceito de povo e de nação em detrimento das especificidades e diferenciações culturais” (KREUTZ, 2000, p. 351). O Brasil é o segundo país da América do sul em contingente imigrante entre os anos de 1819 e 1947: foram cerca de 4.903.991. No sul do país, os imigrantes estabeleciam-se em núcleos etnicamente homogêneos, mas no estado de São Paulo, por exemplo, que recebeu o maior contingente de imigrantes, esses não se organizaram em colônias; foram inseridos na base da pirâmide social e tinham como modelo lingüístico o português popular utilizado pelas camadas populares da sociedade. “Contudo, em função do seu *back-ground* cultural, esses imigrantes ascenderam rapidamente na estrutura social, levando para o seio da norma culta algumas das estruturas de matiz popular que haviam adquirido do seu contato inicial com o português” (LUCCHESI, 2004, p. 109).

A democratização dos padrões culturais e lingüísticos consolidou a tendência de afrouxamento normativo do português culto, esboçada desde a independência do país, quando as manifestações de nacionalismo político e lingüístico buscavam exprimir a ruptura com a ex-metrópole. A expansão da mídia e a popularização dos meios de comunicação como rádio e televisão proporcionavam às camadas mais baixas da sociedade acesso à norma culta, assim como a vulgarização do sistema de ensino também o fez. Até metade do século XX, a grande maioria da população do país era analfabeta.

Todo esse contexto não só aproxima o comportamento lingüístico dos dois pólos como dissolve as marcas mais fortes da aquisição imperfeita do PB na fala das camadas populares da sociedade. Como resultado desse contexto, observa-se duas tendências de mudanças distintas: uma, verificada nas camadas populares, de mudança “para cima”, em direção ao padrão urbano culto; e outra, nas camadas média e alta da sociedade, de mudança “para baixo”, num afastamento do padrão normativo de matiz europeu (LUCCHESI, 2001, p. 109).

Nesse contexto, estão em vias de desaparecimento as marcas mais drásticas das alterações produzidas pelo extenso contato entre línguas ocorrido nos séculos anteriores, mantendo-se os vestígios das características mais expressivas do processo de transmissão lingüística irregular apenas naqueles dialetos rurais mais afastados, em que o contato foi mais radical e a reestruturação lingüística, mais profunda (LUCCHESI, 2001, p. 107).

Dessa forma, as comunidades quilombolas, clandestinas e ilegais ainda representariam uma situação social favorável à manutenção de marcas lingüísticas crioulezantes. Somente a partir de meados de século XX essas regiões viriam a ser exploradas por grandes empreendimentos econômicos que expulsariam os seus habitantes, ou os integraria às linhas de trabalho. Portanto, as características lingüísticas crioulezantes só seriam encontradas em poucas comunidades rurais afro-brasileiras isoladas, as quais estariam, atualmente, passando por um processo de mudança em direção ao padrão urbano propiciado pela vulgarização do sistema de ensino e pela massificação da mídia. Os traços lingüísticos que apontam as mudanças drásticas sofridas pelo PB seriam mais sensíveis na fala dos indivíduos mais velhos dessas comunidades, que, ao contrário de seus descendentes, não tiveram acesso à escolarização ou a meios de comunicação de massa, como o rádio e a televisão.

2 A INDETERMINAÇÃO DO SUJEITO

Muitas já foram as análises feitas sobre a possibilidade de indeterminar e sobre o que seria a indeterminação do sujeito, e muitos foram os esclarecimentos proporcionados por estudiosos como Milanez (1982), Cunha (1993) e Cavalcante (1999), por exemplo; porém poucas foram as alterações nas gramáticas normativas: a possibilidade de pronomes como *a gente*, *você*, *nós* ou *eles* exprimirem a indeterminação ainda é tratada em uma pequena nota, ao fim da página, como se essas estratégias fossem menos legítimas do que as estratégias padrão. Milanez (1982) já atentava para a utilização de formas despronominalizadas – dentre elas a forma \emptyset + 3ª pessoa singular – por falantes da norma culta de São Paulo. Cavalcante (1999) prova que a implementação de estratégias bem distintas das estratégias padrão em nossa estrutura lingüística não é recente: seu estudo observou a imprensa carioca dos séculos XIX e XX. Além de elencar e descrever o comportamento desta variável, estudiosos também discutem largamente questões como o conceito de *determinação* e de *definição*, assim como o comportamento da partícula *se*, sem que, no entanto, as gramáticas normativas se preocupem em ao menos descrever de modo mais fiel a norma do português do Brasil. É inegável que os exemplos como (11) e (12) fornecem semanticamente a noção de um agente indeterminado⁵.

(11) “É assim de... de assalto, essas coisa assim **a gente** fica com medo...”

(12) “Aí eu falei pa ele pedí um atestado po médico pa encostá ele, que aqui encosta.

Eles dão benefício assim.”

A possibilidade de indeterminar fazendo usos de pronomes pessoais e a capacidade de fazê-lo em diferentes graus não é abordada ou reconhecida pelas Gramáticas Tradicionais (doravante GT) como uma possibilidade presente na estrutura da língua portuguesa, assim como também não é considerada a importância do discurso no processo de indeterminação. Em suma, a indeterminação do sujeito está longe de ser plenamente descrita pelas gramáticas

⁵ Exemplos retirados do *Corpus do Português Afro-brasileiro* do Projeto Vertentes da Universidade Federal da Bahia.

pedagógicas de modo que todas as propriedades e minúcias relativas ao assunto sejam reconhecidas.

Serão apresentadas, nos dois primeiros itens deste capítulo, as abordagens de algumas GTs e o tratamento dado ao fenômeno da indeterminação por parte de alguns pesquisadores que abordaram o tema: Milanez (1982), Rollemberg et al (1991), Cunha (1993), Cavalcante (1999), Ilari *et al* (1996), Neves (2000) e Santana (2006). Em seguida, será analisada a propriedade gradativa da indeterminação e a sua abordagem segundo três estudos, bem como a definição do termo *indeterminação* mediante a análise de outros dois fenômenos: *determinação* e *indefinição*. E, por fim, o caso do *se* indeterminador será tema de breve revisão.

2.1 O ENFOQUE DA GRAMÁTICA NORMATIVA

Sabe-se que o número de publicações de gramáticas de língua portuguesa é vasto, e que uma análise crítica dessas e das diferentes abordagens sobre o sujeito indeterminado que seriam encontradas, por si só, seriam tema de um rico estudo. Utilizo-me aqui, com fins de análise da definição e dos recursos de indeterminação, de trechos de apenas algumas gramáticas normativas mais recentes, muito conhecidas de estudantes de Letras: Rocha Lima (1972), Cegalla (1997), Cunha e Cintra (2001) e Bechara (1992).

Rocha Lima (1972, p. 235) aponta duas formas de indeterminação e sua definição de sujeito indeterminado se inicia pela apresentação da noção de sujeito determinado. A distinção entre os dois conceitos é fundamentada na identificação.

O sujeito ainda pode ser determinado, ou indeterminado.

É determinado, se identificável na oração – explícita ou implicitamente; indeterminado, se não pudermos ou não quisermos especificá-lo.

Para indeterminar o sujeito, vale-se a língua de um dos dois expedientes:

- 1) Empregar o verbo na 3ª pessoa do plural, sem referência anterior ao pronome eles ou elas, e a substantivo no plural;
- 2) Usá-lo na 3ª pessoa do singular acompanhado da partícula *se*, desde que o verbo seja intransitivo, ou traga complemento preposicional.

Cegalla (1997, p. 296) soma às duas estratégias de indeterminação clássicas – verbo flexionado em terceira pessoa do plural e verbo intransitivo, de ligação ou transitivo indireto em terceira pessoa do singular acompanhado da partícula *se* – uma terceira: o verbo no infinitivo impessoal. A definição para sujeito indeterminado, porém, é vaga.

- 8) indeterminado – quando não se indica o agente da ação verbal:

Em português assinala-se a indeterminação do sujeito de três modos:

- 1) usando-se o verbo na 3ª pessoa do plural, sem referência a qualquer a gente já expresso nas orações anteriores;
- 2) com um verbo ativo na 3ª pessoa do singular, acompanhado do pronome se;
- 3) deixando o verbo no infinitivo impessoal.

Ex: É triste assistir a estas cenas repulsivas.

Cunha e Cintra (2001, p. 128) definem o sujeito indeterminado como aquele que “não vem expresso na oração e nem pode ser identificado”. As estratégias de indeterminação, assim como em Rocha Lima (1972), são limitadas a apenas duas.

Algumas vezes, 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. Dizemos então, que o sujeito é indeterminado. Nestes casos em que o sujeito não vem expresso na oração nem pode ser identificado, põe-se o verbo:

- a) ou na 3ª pessoa do plural
- b) ou na 3ª pessoa do singular, com o pronome se.

Bechara (1992, p. 200), com relação à conceituação, assim como o que é visto Cegalla e Cunha e Cintra, apresenta o sujeito indeterminado como o agente que não podemos ou não queremos identificar:

Sujeito indeterminado é o que não se nomeia ou por não se querer ou por não se saber fazê-lo.

No que concerne às estratégias de indeterminação apresentadas, o autor apresenta três – agrega às duas estratégias clássicas “o verbo da oração na terceira pessoa do singular sem referência a uma pessoa determinada”.

Quanto à definição do termo indeterminação, apenas Rocha Lima parte da noção de determinação para apresentá-lo. Observando o que se diz nas gramáticas tradicionais acerca do sujeito indeterminado, é possível identificar cinco problemas, tanto na definição e na descrição do fenômeno quanto na metodologia utilizada para tanto:

- (i) a redução do número de estratégias de indeterminação empregadas pelo falante para duas ou três;
- (ii) a falta de clareza no tratamento dos parâmetros semântico e sintático;

(iii) o tratamento que é dado ao fenômeno da indeterminação trata de maneira indistinta dois níveis de referencialidade – um mais genérico, representado, nestas descrições, pelas construções *verbo em 3ª p.s. + se*, e um nível mais específico, nas construções com o verbo em 3ª p.p.;

(iv) A falta de fundamentação semântica no tratamento do fenômeno da indeterminação;

(v) a ausência de reflexões sobre a partícula *se* e sobre o contexto em que esta se insere, visto que, não de hoje, é constatada a não-flexão verbal em construções como *Vende-se casas*.⁶

2.2 ANÁLISE DO TEMA NO ÂMBITO DA LINGÜÍSTICA MODERNA

Diversas análises lingüísticas nas últimas décadas trataram da indeterminação do sujeito em diversos universos de observação no âmbito do português brasileiro. Três análises contemplam a norma culta e utilizaram *corpora* de língua falada. Milanez (1982) verifica a indeterminação do sujeito em entrevistas do Projeto NURC de São Paulo⁷, para isso se vale tanto de construções com sujeitos pronominais quanto com sujeitos lexicais. Rollemberg *et al* (1991) utiliza o *corpus* do NURC constituído na cidade de Salvador e verifica o comportamento das formas verbais com sujeito realizado e formas verbais com sujeito lexical zero. Cunha (1993) estuda o comportamento dos pronomes *a gente*, *nós* e *você* indeterminados na norma urbana culta do Rio de Janeiro. A indeterminação do sujeito em língua escrita é o foco do trabalho de Cavalcante (1999). O *corpus* estudado em Cavalcante foi composto por escritos da imprensa carioca dos séculos XIX e XX e as estratégias analisadas distribuem-se por dois grupos: as estratégias com formas verbais finitas e as estratégias com formas verbais não-finitas. Finalmente, completando o conjunto de trabalhos aqui revistos, temos trabalhos que integram duas gramáticas: Ilari *et al* (1996), em capítulo dedicado aos pronomes pessoais no português, na *Gramática do Português Falado*, descrevem o comportamento desses na norma urbana culta⁸ e, diferentemente do que é visto nas gramáticas tradicionais, ressaltam também a função indeterminadora; e Neves (2000), na *Gramática de Usos do Português*, que

⁶ Essa questão sera abordada mais adiante, no item 2.5

⁷ *Corpus* constituído no **Projeto de Estudo da Norma Lingüística Urbana Culta no Brasil**.

⁸ Ilari *et al* utilizam o *corpus* mínimo do NURC.

aborda também a capacidade generalizadora dos pronomes pessoais. Mais recentemente, Santana (2006) tratou a indeterminação do sujeito no português rural do semi-árido baiano. Em sua dissertação de Mestrado, foram descritos os comportamentos de estratégias finitas e não-finitas do vernáculo das comunidades de Barra e Bananal, no município de Rio de Contas, também estudadas neste trabalho.

Os exemplos (13) a (18), apresentados a seguir, exemplificam as ocorrências de indeterminação do sujeito em orações finitas levantadas por alguns dos pesquisadores já citados⁹. Nelas figuram, nesta ordem, as seguintes formas de indeterminação do sujeito, os pronomes *a gente*, *você*, *nós* e *eles*, e as formas verbais da 3ª pessoa do singular e do plural sem sujeito expressos (aqui referidas, respectivamente, por $\emptyset + V3PS$ e $\emptyset + V3PP$).

(13) A caipirinha *a gente* prepara com limão galego.

(14) Se *você* pede 10 mil cruzeiros emprestado, o banco desconta pelo menos uns 5%...

(15) A Caixa Econômica, *você* vai tomar dinheiro emprestado na Caixa Econômica, normalmente *eles* exigem uma garantia hipotecária.

(16) Eu acho que *nós* estamos caminhando pra uma, uma situação muito séria de publicidade, ouviu? Por que *nós* estamos preparando as criaturas pra ser dominadas mentalmente pela publicidade.

(17) O piano é o tipo de instrumento que resulta imediatamente, por que o som está pronto, né? *Chega* lá, *bota* a mão. Está tudo no lugar.

(18) A comunhão na mão eu acho que é só pra... pra equipe de casais, de pessoas que já... eu acho... que... estão mais ao par, por que em Lourdes não me *deram* na mão por que *perceberam* que eu estranhei.

Em sua análise da indeterminação do sujeito na norma urbana culta de São Paulo, capital, com base nos materiais do projeto NURC, Milanez (1982) analisou a fala de 20 infor-

⁹ O exemplo (13) foi retirado de Cunha (1993), os exemplos restantes são de Milanez (1982).

mantes, em entrevistas que totalizaram vinte horas. O *se* foi considerado indeterminador tanto em construções com verbos intransitivos, transitivos indiretos e de ligação como em construções com verbos transitivos diretos, em que a partícula *se* obteve a maior frequência - 87%. As estratégias de indeterminação, neste trabalho, foram classificadas em dois grupos: indeterminação com sujeito lexical, em que constam ocorrências dos pronomes *a gente*, *você* e *eles*; e a indeterminação sem sujeito lexical, forma a que pertencem as construções (\emptyset + 3ª p.s.), (\emptyset + 3ª p.p.), (\emptyset + verbo + se) e (\emptyset + infinitivo)¹⁰. Além de estratégias pronominais, Milanez também levantou ocorrências de estratégias lexicais como *indivíduo*, *sujeito*, *pessoa*, entre outras. De 1058 ocorrências quantificadas, 62% destas eram de sujeitos realizados e somente 38% foram de formas despronominalizadas. O pronome *a gente* obteve o maior número de ocorrências, seguido da partícula *se* e do pronome *você*. Os resultados apresentados em Milanez (1982) podem ser vistos no quadro a seguir.

Tabela 3 – Levantamento Geral em Milanez (1982)

	ESTRATÉGIA	QUANT.
1º	A gente	361
2º	Se	205
3º	Você	168
4º	\emptyset + 3ª p.s.	120
5º	\emptyset + 3ª p.p.	53
6º	a pessoa	45
7º	\emptyset + infinitivo	30
8º	o indivíduo	28
9º	o sujeito	16
10º	Eles	13
11º	Eu	11
12º	o cara	07
	TOTAL	1058

No entanto, ao se excluir destes resultados as formas lexicais e a forma não finita (\emptyset + infinitivo), tem-se o seguinte resultado:

¹⁰ A nomenclatura que aqui se apresenta é a utilizada pela autora.

Tabela 4 – Levantamento Geral em Milanez (1982) adaptada

ESTRATÉGIA	QUANT.	%
A gente	361	38,8
Se	205	22
Você	168	18
Ø + 3ª p.s.	120	12,9
Ø + 3ª p.p.	53	5,7
Eles	13	1,4
Eu	11	1,2
TOTAL	931	100%

Os pronomes pessoais em função indeterminadora na norma urbana culta da cidade de Salvador são o alvo do estudo de Rollemberg *et al* (1991). Foram levantadas as ocorrências dos pronomes sujeito *você*, *nós*, *a gente* e *eles*, focalizados no estudo, em quatro inquéritos que compõem o “*corpus compartilhado*” do NURC/SSA¹¹. As ocorrências das formas verbais despronominalizadas ($\emptyset + 3^a$ p.s. + *se*), ($\emptyset + 3^a$ p.p.), ($\emptyset + 3^a$ p.s.) e ($\emptyset +$ infinitivo) não foram registradas. Os índices gerais obtidos pelas estratégias pronominais são apresentados na tabela 5.

Tabela 5 – Frequência das estratégias de indeterminação do sujeito em Rollemberg (1991).

ESTRATÉGIA	QUANT.	%
Você	75	35,54
Nós	68	32,23
A gente	53	25,12
Eles	15	7,11
TOTAL	211	100

Além da variável dependente, foram consideradas no estudo uma variável discursiva e uma variável social: a categoria do texto, cujas variantes são elocução formal e diálogo entre informante e documentador, e faixa etária. Foram examinadas somente as duas faixas extremas: faixas I, de 25 a 35 anos, e faixa III, de 56 anos em diante. Na faixa I, prevalece o uso

¹¹ O corpus compartilhado é composto por noventa inquéritos, dezoito de cada uma das cinco cidades estudadas pelo projeto NURC.

da forma *você*, enquanto o pronome *eles* é a estratégia mais favorecida pela faixa III (cf. quadro 1).

Quadro 1 – Os pronomes pessoais na indeterminação do sujeito e a faixa etária.

PRONOMES	FAIXA I	FAIXA III
Você	75 (100%)	-
Nós	35 (51,74%)	33 (48,53%)
A gente	43 (81,13%)	10 (18,87%)
Eles	02 (13,33%)	13 (86,67%)
TOTAIS	155 (73,46%)	56 (26,54%)

Quadro extraído de Rollemberg (1991)

Cunha (1993) analisou o comportamento das estratégias de primeira e segunda pessoas e os fatores que influenciariam a escolha por uma ou outra estratégia de indeterminação. As ocorrências foram distribuídas segundo oito estratégias, mas, para a análise, o *corpus* foi reduzido às ocorrências de três pronomes, *nós*, *a gente* e *você*. Os índices obtidos foram de 48% para *você*, 29% para *nós* e 23% para *a gente*. A forma antecedente foi a variável considerada pelo pacote de programas VARBRUL¹² como mais relevante do ponto de vista estatístico para os três pronomes em questão. A distribuição das estratégias pelas variáveis sociais sexo e faixa etária apontou resultados interessantes: “*nós* e *a gente* são as formas preferidas pelas mulheres e *você* é a forma preferida pelos homens” e “as mulheres lideram a mudança de *nós* por *a gente*, quando, nas faixas etárias mais jovens e em elocuições menos formais, fazem uso desta última forma para indeterminar seus enunciados”. (CUNHA, 1993, p. 101)

A utilização e, ainda, a implementação em nossa estrutura lingüística de estratégias outras que não somente as consideradas padrão não são fenômenos recentes. Cavalcante (1999) analisou o comportamento do sujeito indeterminado em textos da imprensa carioca em

¹² Para mais explicações, ver o item 3.2.3, no capítulo seguinte.

cinco períodos, que vão do ano de 1848 ao ano de 1998¹³, e constatou o uso das formas *nós*, *eles*, *a gente* e *você* com valor indeterminador, além das estratégias \emptyset + verbo em 3ª p.p., \emptyset + verbo em 1ª p.p., \emptyset + verbo + *se*, verbo no infinitivo com e sem a partícula *se* e verbo no gerúndio. Quanto às formas verbais finitas, que totalizaram 2.355 ocorrências, verificou-se cinco fatos importantes:

- a) a diminuição do uso do pronome *nós* no decorrer do período estudado;
- b) o aumento do uso da partícula *se*;
- c) a baixa frequência da estratégia padrão \emptyset + verbo em 3ª p.p., que obteve, respectivamente, 7%, 16%, 2%, 7%, e 8% de frequência ao longo dos cinco períodos analisados;
- d) a utilização da estratégia *a gente* já no primeiro período analisado no estudo, que vai de 1848 a 1869;
- e) a recente implementação da estratégia *você*, que só aparece a partir do período compreendido entre os anos de 1961 e 1968.

As ocorrências de sujeito preenchido não ultrapassaram 9% do total de ocorrências e foram mais frequentes em construções com os pronomes *a gente* e *você*, em que o verbo não possui marcas verbais específicas.

Mais recentemente, Neves (2000), na *Gramática de Usos do Português*, analisa algumas estratégias de indeterminação. A subseção específica para o comportamento dos pronomes indeterminados intitula-se *Os pronomes pessoais podem fazer referência genérica*, e nesta são descritos os comportamentos dos pronomes *você*, *eu*, *eles* e *nós*, quando indeterminados, e também são analisadas estratégias como \emptyset + verbo em 1ª p.p., \emptyset + verbo em 3ª p.s. e as construções com partícula *se*. Com relação à esta última estratégia, ainda são mantidas distinções entre o seu valor indeterminador e apassivador.

Também no volume quatro da *Gramática do Português Falado*, organizada por Ataliba Teixeira de Castro e Margarida Basílio, é descrito o comportamento dos pronomes em função indeterminadora. A pesquisa foi realizada por Rodolfo Ilari, Carlos Franchi, Maria Helena Moura Neves e Sírío Possenti, e o *corpus* utilizado foi o “*corpus mínimo*” do NURC, constituído de quinze inquéritos, três por cidade, a saber São Paulo, Rio de Janeiro, Recife, Porto Alegre e Salvador. O capítulo intitula-se *Os pronomes pessoais no português falado*:

¹³ Os cinco períodos estudados em Cavalcante são: Período I, de 1848 a 1869; período II, de 1891 a 1910; período III, de 1935 a 1942; período IV, de 1954 a 1968; e período V, de 1996 a 1998.

roteiro para a análise e, no item 1.3 – *A referenciação indeterminada*, são feitas as seguintes considerações:

- a) o pronome *nós* apresenta um caráter [+definido] enquanto o pronome *a gente* [+indeterminado];
- b) o pronome *você* exprime um grau máximo de generalização e os pronomes *você*, *nós* e *eu* podem alternar-se;
- c) quando são utilizados os pronomes de terceira pessoa, a generalização não é total;
- d) uma generalização maior parece ser obtida com o uso do *se* junto a alguns verbos;
- e) a referência generalizada com a terceira pessoa do singular – \emptyset + 3ª p.s. – pode sugerir uma interpretação de referência como a de um sujeito oracional; e
- f) também é estratégia de indeterminação o uso de sintagmas nominais com substantivos de sentido genérico, como *o pessoal*, por exemplo.

Quanto à partícula *se*, é importante salientar que na pesquisa apresentada na *Gramática do Português Falado* somente um único falante formulou sentenças passivas, ou seja, verbos que concordavam com SN no plural; as demais construções foram analisadas como indeterminadoras.

Em sua pesquisa sobre a indeterminação do sujeito no vernáculo das comunidades de Barra e Bananal, no município de Rio de Contas, interior do estado da Bahia, Santana (2006) levantou e tratou quantitativamente as ocorrências dos pronomes *a gente*, *nós*, *você* e *eles* e das formas verbais \emptyset + V3PS, \emptyset + V3PP, \emptyset + VINF e \emptyset + V + SE. As frequências alcançadas pelas variantes podem ser conferidas na tabela 6.

Tabela 6 – Frequência dos recursos de indeterminação do sujeito nos corpora

VARIANTES	N	%
A gente	1.149	45
Ø + V3PS	525	20
Eles	272	10,5
Nós	227	9
Você	151	6
Ø + V3PP	120	5
Ø + VINF	109	4
Ø + V + SE	15	0,5
TOTAL	2569	100

Tabela extraída de Santana (2006, p. 93)

Quanto às variáveis lingüísticas explanatórias examinadas por Santana (2006), a forma antecedente mostrou-se como a mais significativa para todas as estratégias de indeterminação, com exceção da variante $\emptyset + V3PP$. No plano do encaixamento social, as variáveis gênero e faixa etária foram selecionadas pelas formas *a gente*, *você* e $\emptyset + V3PS$. Para a forma $\emptyset + VINF$ apenas o gênero mostrou-se relevante. A localidade foi significativa apenas para as variantes *a gente*, *nós* e $\emptyset + V3PP$.

2.3 OS GRAUS DE INDETERMINAÇÃO DO SUJEITO

Sabe-se que a indeterminação não se opõe claramente à determinação por ser um fenômeno de natureza gradual. As informações sobre a referência de cada ocorrência são obtidas ao longo da fala do informante, sendo necessária, portanto, uma análise de todo o discurso para que seja confirmada a indeterminação. Exemplos desta gradação podem ser observados a seguir.

Milanez (1982), sobre esta questão, afirma que a indeterminação comporta três graus quanto à sua abrangência. O grau mais alto é expresso pelas formas despronominalizadas ($\emptyset + 3^a$ p.s., $\emptyset + 3^a$ p.s. + se e $\emptyset +$ infinitivo¹⁴), que, como não possuem marcar de pessoa, distanciam-se das três pessoas do discurso. Um grau intermediário seria de efeito não tão genérico quanto o grau mais alto de indeterminação – ainda manteria referências específicas aos protagonistas do diálogo. Este grau seria expresso por formas como *a gente*, *você* e *eu*, por

¹⁴ A nomenclatura é a utilizada pela autora.

exemplo. O menor grau da indeterminação é representado pelas formas $\emptyset + 3^a$ p.p. e *eles*, exclusivas de terceira pessoa e que normalmente ocorrem anteceditos por uma referência a um grupo social específico. Quanto à identificação dos graus de indeterminação, a autora afirma que

dois fatores foram relevantes: a própria forma lexical dos recursos, já que determinada, de certa forma, a amplitude de generalização dos mesmos, e o contexto, que desempenha, muitas vezes um papel “controlador” dessa abrangência. (p. 82)

Desta maneira, é admitido que as estratégias que atendem aos graus de indeterminação são classificadas em graus, fazendo parte cada uma de grupos distintos.

Para Cunha (1993), a indeterminação do sujeito possui “uma escala, em que a indeterminação cresce à medida que deixam de existir elementos no contexto que permitam a interpretação do pronome”. Haveria então três graus de indeterminação: *indeterminação parcial com elo de referência explícita*, em que os pronomes relacionam-se com um item lexical do texto; *indeterminação parcial com elo de referência implícita*, em que a referência do pronome é depreendida pelo contexto sem que haja um item lexical presente; e a *indeterminação completa*, em que não há relação com nenhum item ou relação anafórica. A indeterminação parcial é representada principalmente pelo pronome *nós*, que é responsável por 67% das ocorrências nesse grau. O pronome *a gente* corresponde a 30% dos enunciados e o *você*, a apenas 3%. Quanto à indeterminação completa, observa-se o inverso: o pronome *você* alcança a frequência de 72%, seguido do pronome *a gente*, com 19%, e do pronome *nós*, com 9%.

Santana (2006) utiliza-se da gradação proposta por Cunha (1993). O grau de indeterminação, na análise quantitativa dos dados, foi selecionado por todas as formas indeterminadoras estudadas. Segundo Santana,

Nós e *a gente* parecem ter o mesmo valor: são sempre usadas para indicar indeterminação parcial com referência implícita ou explícita no contexto; *eles* é sempre usado na indeterminação parcial com referência explícita no contexto; e *você*, $\emptyset + V3PS$, $\emptyset + V3PP$ e $\emptyset + VINF$ na indeterminação completa, por terem uma abrangência genérica, podendo referir-se a qualquer pessoa do discurso. (p. 140)

A variável *nível de refencialidade do sujeito*, em Ponte (2006), utilizada na presente pesquisa, é abordada como possuidora de quatro níveis: *genérico*, em que a pessoa utilizada pelo falante engloba qualquer pessoa do discurso; *parcialmente definido*, em que o sujeito sem referência anterior explícita refere-se a um grupo sem contornos definidos, mas implícito

do discurso; *específico indefinido*, em que o praticante da ação não se pode ou não se quer determinar; e um nível mais específico, o do *Eu ampliado*, representativo de casos em que há alternância dos pronomes *nós* e/ou *a gente* e o pronome de 1ª pessoa *eu* e em que se verificam a utilização de *nós* e *a gente* com referência à primeira pessoa do singular. Este último nível tem como base Benveniste (1995 [1946]). Sabe-se que os pronomes pessoais de primeira e segunda pessoas, em singular, fazem referência a elementos específicos do discurso e não são pluralizados, pois são únicos e não permitem quantificação: o *nós* não é (*eu + eu*), mas (*eu + outras pessoa*). Segundo a autora:

Esse “nós” é algo diferente de uma junção de elementos definíveis; a predominância do “eu” aí é muito forte, a tal ponto que, em certas condições, esse plural pode substituir o singular. A razão está em que o “nós” não é um “eu” quantificado ou multiplicado, é um “eu” dilatado além da pessoa estrita, ao mesmo tempo acrescido e de contornos vagos. [...] De maneira geral, a pessoa verbal no plural exprime uma pessoa amplificada e difusa. O “nós” anexa ao “eu” uma globalidade indistinta de outras pessoas. (p. 257)

Desta forma, o falante, ao utilizar os pronomes de primeira pessoa *nós* e *a gente*, pode fazer referência a um grupo de contornos vagos, composto por ele mesmo e demais pessoas, ou pode, quando o plural substitui o singular, referir-se apenas a si mesmo. Assim, construções como (19) e (20), que pareciam ser classificadas como pertencentes a apenas um grau de indeterminação, são classificadas como pertencentes a dois graus distintos: *o grupo parcialmente definido* e *o eu ampliado*.

(19) Quato lata *a gente* vende... aqui na roça vende... até aqui na roça também!

(20) Aí, pronto, aí, *a gente* ajudava mãe, quando *a gente* era pequena... mas pequena assim não, porque... **pequena** eu digo assim... já tinha mais de dez ano, quando **eu** comecei a fazê charuto, sabe?

No exemplo (19), o falante comenta uma venda praticada pelas pessoas da sua comunidade, trata-se, portanto, de uma ocorrência de indeterminação em *grupo parcialmente definido*. Já no exemplo (20), percebe-se que o pronome *a gente* é utilizado pela falante para referir-se a si mesma. Esta interpretação é confirmada aí pelo uso da concordância em “*a gente* era *pequena*” e pelo uso, em seguida, do pronome *eu*. O falante amplia a abrangência do

seu discurso ao utilizar a primeira pessoa do plural, portanto essas ocorrências foram inseridas no nosso universo de observação.

A seguir, são apresentados e discutidos exemplos dos níveis de referencialidade propostos por Ponte (2006) e utilizados aqui.

(21) É, *a gente* procura onde tá a terra mais molhada, *a gente* sabe que dá água mais rápido, né?

(22) Se *você* ‘tivê com dô, aí *você* tem que tomá qualqué remédinho em casa mesmo.

(23) Aí depois cabá de rapa [a cana de açúcar], agora Ø rela... agora Ø põe na prensa. No ôto dia é que Ø vai fazê.

(24) As coisa mais violenta, aí começô mais passá moto, começô mais aquela coisa... aquelas estada, abrir mais estrada aqui pra Santo Antônio e tudo, qué dizê, que aí vai os malando sai até pelos mato... dento dos mato e aí *a gente* fica com medo.

(25) Num tinha esse negócio de ‘tá dano mal criação ao mais velho... Quando o mais velho falava, *nós* tinha que obedecê...

(26) DOC: E... aí, vocês comemora o natal também, né?

INF: Não. Nós não comemoramos o natal nem festa junina, né?

DOC: É, junina sei que não. E o natal também não?

INF: Não. Por que Ø num comemora.

(27) INF: Quem? Eu acho que é pelo governo esses banheiros. Acho que é pelo governo que...

DOC2: É o quê? Faz o cadastro...

INF: É. Aí *eles* da aí assim minha senhora, Ø dá os meterial e aí Ø dá as pessoa, assim, pra construi(r) aqueles banheiro...

(28) Num tenho esperança de nada! *Tomaro* o nome, *fizero* eu preencher um... uma ficha... esqueci o nome do... da ficha... e aí pronto!

(29) ‘Tavam discutino sobre esse negoço desse... dessa turma que diz que Ø chega assim e que Ø ‘tá matano todo mundo!’

Os três primeiros exemplos apresentados, (21), (22) e (23), apesar de terem sujeitos distintos – respectivamente, *a gente*, *você* e a categoria vazia – são classificados como pertencentes a um grau maior de indeterminação, de caráter mais genérico. Os falantes, em seus discursos, referem-se a processos dos quais qualquer pessoa poderia ser o sujeito: como buscar um local para abrir um poço, no exemplo (21); o modo como se procede quando não há um posto de saúde em sua comunidade, no exemplo (22); e o processo de tratamento da cana de açúcar, no exemplo (23). Todas as pessoas, uma vez envolvidas nesses processos, devem agir do mesmo modo, e não há restrições quanto à um grupo que desempenha estas atividades.

Os três exemplos seguintes, (24), (25) e (26), ao contrário do que foi visto anteriormente, já possuem limites quanto ao agente da ação, contudo esses limites não são precisos. No exemplo (24), a noção de grupo é alcançada através localidades citadas pelo falante, como “a estrada aqui pra Santo Antonio”. Entende-se, portanto, que quem “fica com medo” são as pessoas que moram pelas proximidades. No exemplo (25), cujo sujeito é o pronome *nós*, a referência não é explícita, mas é obtida pelo contexto: os mais novos, grupo no qual o falante se inclui, deveriam obedecer se os mais velhos falassem. No exemplo (26), a falante refere-se ao grupo composto por ela e as demais pessoas que seguem a mesma religião. O grupo, inicialmente representado pelo pronome *nós*, passa, após o turno do documentador, a ser representado pela categoria vazia, entretanto a noção de grupo, presente no contexto, permanece.

Um terceiro grau de indeterminação é o do sujeito específico, cuja referência não é relevante ou não nos foi dado conhecer, a que as gramáticas tradicionais exemplificam com orações com verbo em 3ª pessoa do plural. Entretanto este grau de indeterminação também pode ser expresso por outras formas. No exemplo (27), o sujeito *elas* faz referência a pessoas específicas que doam aos moradores material de construção. A mesma noção é desempenhada pelo sujeito do exemplo (28) – construção em que figura o verbo flexionado na terceira pessoa do plural sem sujeito realizado – e pela categoria vazia no exemplo (29). Em todas as orações, o sujeito são pessoas específicas, porém não reveladas.

2.4 DETERMINAÇÃO, INDETERMINAÇÃO E INDEFINIÇÃO

2.4.1 Indeterminação e determinação

A delimitação do conceito de *indeterminação*, nos estudos tradicionalistas, ainda hoje é objeto de discussão por parte dos estudiosos da área; não raro o conceito de *indeterminação* é confundido com o de *indefinição* ou de *impessoalização*. Agrega-se também a este problema descritivo um metodológico: as GTs raramente partem do conceito de *determinação* para chegar ao de *indeterminação*. Contudo muitos estudiosos, como Milanez (1982) e Cunha (1993), por exemplo, proporcionaram-nos diversos esclarecimentos.

Segundo Milanez, para a análise do conceito de *indeterminação*, torna-se necessário que consideremos as noções de pessoa, generalização e referência. A determinação se dá quando as formas verbais ligam-se a seres específicos e identificáveis no discurso; a *indeterminação*, portanto, é “o processo da língua que permite ao falante passar do universo das três pessoas identificáveis [...] a um nível de generalização”, em que a referência da pessoa do discurso é abrangente. Esta oposição entre sujeito determinado e indeterminado pode ser facilmente identificada quando se analisa os exemplos a seguir.¹⁵

(30) Você gostou do jantar? (2ª pessoa)

(31) O gato pulou o muro. (3ª pessoa)

(32) O meu irmão e eu não gostamos de festa. (1ª e 3ª pessoas)

(33) No caso *nós* tamo em dois mil e três... dois... dois mil e dois, né?

Nos exemplos (30), (31) e (32), as construções “fazem referência a uma ou mais pessoas do discurso de forma nítida e evidente” (MILANEZ, 1982). *Você*, *o gato*, e *meu irmão e eu* são os sujeitos das orações apresentadas e esta identificação se faz de forma clara. No exemplo (33), porém, o pronome *nós* refere-se, não a um grupo composto pela junção da 1ª

¹⁵ Os exemplo (30) - (32) foram retirados de Milanez (1982). O exemplo (33) pertence ao *corpus* utilizado nesta pesquisa.

pessoa e outrem, mas refere-se a todas as pessoas. O falante atribui ao pronome um caráter genérico, que não lhe permite referência específica.

Quanto à indeterminação pronominal, Cunha (1993) trata como determinados “somente os pronomes que mantenham com seu referente um elo semântico, pragmático e sintático plenamente identificável e recuperável, levando a apenas uma possibilidade interpretativa”. Por indeterminados, a autora toma “aqueles [pronomes] cuja referência não possa ser diretamente recuperada” (p. 34).

2.4.2 Indeterminação e indefinição

O conceito de *indefinição*, segundo Milanez (1982), distingue-se do conceito de *indeterminação* pelos seguintes aspectos:

- a) a indefinição não abrange somente elementos humanos, característica própria da indeterminação;
- b) a indefinição abrange com exclusividade as formas de 3ª pessoa, a indeterminação pode abranger a 1ª, 2ª, ou 3ª pessoas;
- c) a generalização é uma possibilidade de natureza limitada na indefinição, mas é uma condição da indeterminação;
- d) a indefinição apresenta distinções morfológicas relevantes, como o uso dos artigos definidos ou indefinidos, que não se apresenta na indeterminação;
- e) a interpretação da indefinição não depende do contexto, como a da indeterminação, e pode ser alcançada na frase;
- f) o processo de indefinição pertence ao nível de determinação, pois se identifica de modo preciso as pessoas envolvidas no mesmo;
- g) a indefinição é um processo abrangido pela esfera da não-pessoa, já que se limita a 3ª pessoa do discurso.

2.5 A PARTÍCULA SE

Segundo Naro (1976) e Nunes (1991), a função original da partícula *se* seria a função reflexiva. Com a diversificação dos usos do *se*, a passividade passou também a ser represen-

tada por ele. Nunes (1991) levanta a hipótese de que o surgimento do *se* indeterminador tenha sido propiciado por um processo de reanálise semântica nas construções com o *se* apassivador (34).

(34) Ø Consertou-se o brinquedo.

Para a GT, as construções com o *se* apassivador e com o indeterminador se distinguem pela transitividade do verbo. Assim, os exemplos (35) e (36)¹⁶ seriam classificados distintamente somente pela presença do objeto direto *charuto*. Enquanto o verbo *fumar* em (35) seria, por não apresentar objeto direto, classificado como verbo intransitivo e, consequentemente, pertenceria a uma oração com sujeito indeterminado, o mesmo verbo em (36) seria classificado como transitivo direto e, portanto, pertenceria a uma oração passiva.

(35) Fuma-se aqui.

(36) Fuma-se charuto aqui.

Contudo, já há algum tempo, estudiosos têm defendido uma nova posição perante a interpretação destes exemplos.

Milanez (1982) analisa as ocorrências da partícula *se*, seja qual for a transitividade do verbo a que a partícula acompanha, como uma ocorrência de indeterminação e vale-se para isto de quatro argumentos:

1º) a intuição do falante de que a construção remete a um sujeito indeterminado;

2º) o argumento de que deveria haver um só critério para a identificação do sujeito indeterminado, e o que normalmente se vê são dois critérios para um mesmo fenômeno: um critério semântico – na sentença *Fuma-se aqui*, o vazio lexical é interpretado como elemento semântico que “não podemos ou não queremos nomear” – e um critério sintático – a classificação da oração *Fuma-se charutos aqui* como uma oração passiva baseia-se na concordância;

3º) o fato de as orações que a gramática normativa classifica como passiva sintética e passiva analítica não levarem o falante a uma mesma interpretação.

¹⁶ Exemplo retirado de Milanez (1982).

4º) o *se* “índice de indeterminação do sujeito” e o *se* “partícula apassivadora” têm traços comuns, como ambos poderem acompanhar verbos cujos sujeitos têm ou exigem o traço (+ humano).

Embora se tenha obtido, no estudo que se segue, poucas ocorrências com a partícula *se*, não foram feitas distinções quanto à transitividade verbal da mesma.

3 FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA

3.1 A TEORIA

3.1.1 A Pesquisa Sociolingüística

A Sociolingüística Variacionista Quantitativa, ou Laboviana, inicia-se durante a década de 60 e tem como tarefa determinar a sistematicidade da variação. A partir dos estudos realizados por William Labov, sob a orientação Uriel Weinreich, acerca da variação na centralização do núcleo dos ditongos (ay) e (aw) na fala dos habitantes da ilha de Martha's Vineyard, tem início um novo modelo teórico-metodológico, no qual a língua deixa de ser vista como estrutura estática e que tem como objeto de estudo a mudança. Neste estudo, Labov relaciona fatores sociais, tais como sexo, origem étnica e atitude perante variáveis lingüísticas ao comportamento lingüístico dos habitantes de Martha's Vineyard. Posteriormente, as técnicas utilizadas foram mais desenvolvidas e aplicadas pelo pesquisador numa pesquisa sobre as mudanças fônicas em progresso no inglês de Nova York.

Os estudos da mudança em progresso realizados por Labov convergem para uma nova teoria da mudança lingüística, formalizada entre 1966 e 1968 no *Empirical Foundations for a Theory of Language Change*, de Uriel Weinreich, Willian Labov e Marvin Herzog. A partir da sua apresentação no Simpósio “Direções para a Lingüística Histórica”, em 29 e 30 de abril de 1966, na Universidade do Texas, os autores desenvolvem um novo modelo teórico adequado à questão da mudança lingüística. Para que esta fosse resolvida, a sociolingüística variacionista propunha a resolução de cinco problemas, reunidos e sistematizados em Weinreich, Labov e Herzog (2006 [1968]):

- a) **o problema das restrições:** determinar o conjunto de mudanças e condições possíveis para esta;
- b) **o problema da transição:** definir e analisar o percurso através do qual cada mudança se realiza;
- c) **o problema do encaixamento:** conceber a mudança dentro das estruturas lingüística e social;
- d) **o problema da avaliação:** forma como um falante se posiciona perante á mudança lingüística;

e) **o problema da implementação**: definir a forma como uma mudança é implementada.

Além de entender a variação como princípio geral e universal das línguas naturais humanas, que pode ser descrito e analisado, a sociolinguística parte do pressuposto de que a escolha por uma das formas disponíveis para o falante no seio de uma comunidade de fala, constantemente em coocorrência e concorrência, seria influenciada diretamente por fatores estruturais (ou internos à língua) e sociais (ou externos). Desta forma, um fenômeno variável (ou, simplesmente, uma variável), está sujeito à influência de um grupo de fatores estruturais e sociais, portanto a denominação *variável dependente*, largamente utilizada. Contudo a variação não implica inevitavelmente em mudança linguística, visto que duas ou mais formas podem coexistir indefinidamente na língua; toda mudança, porém, se origina na variação.

Através do estudo da variação no presente, considerando a diferença no comportamento linguístico dos falantes em função da sua faixa etária, tornou-se possível verificar os processos de implementação de uma mudança linguística em curso na estrutura da língua. Esta metodologia auxiliou a superação da dicotomia saussuriana entre *sincronia* e *diacronia*, e, assim, hoje se pode optar por duas formas para analisar a mudança linguística: através de estudos em tempo real ou de estudos em tempo aparente.

Estudos em tempo real são realizados quando se observa o comportamento de uma variável ao longo do tempo considerando a documentação histórica disponível ou quando é feito um estudo em uma dada comunidade de fala e, após um determinado espaço de tempo, retorna-se a esta comunidade para que seja dada continuidade a este estudo. Quanto aos estudos de língua falada em tempo real, devido à tecnologia para a captação de áudio não ser suficiente até depois de meados do século passado, esses contemplam apenas duas gerações de falantes, pois ficam limitados às análises de entrevistas realizadas a partir da década de 70.

O desenvolvimento de uma pesquisa em tempo aparente se dá através de um recorte sincrônico no tempo. Estes estudos baseiam-se na hipótese clássica de que o processo de aquisição da linguagem se encerraria por volta da puberdade. A partir de então, pouco seria modificado no sistema linguístico de um indivíduo, visto que este não teria mais acesso aos dispositivos da linguagem que permitiriam modificar a estrutura da língua adquirida. Segundo Chambers e Trudgill (1980, p.165),

a validade do [tempo aparente] depende crucialmente da hipótese de que a fala das pessoas de 40 anos hoje reflete diretamente a fala das pessoas de 20 anos há 20 anos atrás e pode, portanto, ser comparada com a fala das pessoas de 20 anos de hoje, para uma pesquisa da difusão da mudança lingüística. As discrepâncias entre a fala das pessoas de 40 e 20 anos são atribuídas ao progresso da inovação lingüística nos vinte anos que separam os dois grupos.

Dentre os fatores sociais que podem influenciar a escolha por uma ou outra variante, tem-se o sexo, a idade e a origem do falante, por exemplo, fatores que podem ser controlados mais facilmente que fatores como escolaridade, de limites bem menos claros.

Adiante, na apresentação das variáveis sociais analisadas na presente pesquisa, serão feitas algumas considerações acerca destas.

3.1.2 A Transmissão lingüística irregular e o português popular do Brasil

Processos históricos são altamente variáveis em termos de seus parâmetros sócio-demográficos e etno-lingüísticos, assim é impossível se pensar os processos de pidginização e crioulização como fruto de um único contexto. O processo de reestruturação original da gramática (que ocorre após a simplificação/redução da estrutura gramatical da língua alvo), a depender do histórico particular de cada sociedade, se dá de forma mais ou menos intensa, podendo resultar em uma língua historicamente nova ou numa nova variedade da língua do grupo dominador.

A transmissão lingüística regular, ao contrário do postulado por Naro e Scherre, não seria o processo de aquisição de língua que se dá somente entre “crianças a partir da fase de socialização na base de uma amostra de fala susceptível de uma análise ordenada” (2003b, p. 286), mas sim todo processo de aquisição de língua (materna ou não) que se dá de maneira regular, ou seja, todo processo em que o falante tem acesso não restrito à língua alvo e dispõem de ações normatizadoras que possam balizar o seu aprendizado. O conceito de *transmissão lingüística irregular*, portanto, designa “processos históricos de contato massivo entre línguas nos quais a língua do segmento que detém o poder político é tomada como modelo ou referência para os demais segmentos” e “constitui um contínuo de níveis diferenciados de socialização/nativização de uma língua segunda” (Lucchesi, 2003a, p.272).

O processo a que pertencem estas variedades caracteriza-se fundamentalmente por (LUCCHESI, 2003a, p.276):

- (1) perda ou variação no uso de morfologia verbal e palavras gramaticais;

- (2) alteração dos valores dos parâmetros sintáticos em função de valores não-marcados;
- (3) gramaticalização de itens lexicais para preencher lacunas na estrutura lingüística.

A história do Brasil se encaixa perfeitamente no contexto sócio-histórico que propicia o processo de contato entre línguas de tipo leve. Quando a ocupação efetiva do Brasil, em 1532, se deu, aqui viviam pouco mais de um milhão de indígenas, distribuídos por mais ou menos trezentas línguas diferentes; apesar do genocídio cometido pelos colonizadores, até o século XVIII a situação no país era de bilingüismo. Do período que foi de 1538 a 1855, o Brasil recebeu cerca de 18 milhões de escravos africanos, em sua maioria de duas culturas (Bantos e Sudaneses) que, mais tarde, se integraram à grande massa rural do país, portanto as marcas mais profundas da aquisição defectiva do português se mantiveram somente nas comunidades rurais isoladas. Além disso, o acesso à língua alvo no Brasil foi maior do que o verificado em situações mais típicas de criouliização, pois a população branca no país nunca foi inferior a 30%. (CASTILHO, 1992)

Com base nesta argumentação, acredita-se que muitas das características estruturais do português popular do Brasil resultaram dos processos de transmissão lingüística irregular, dentre elas:

- (1) Variação na concordância de número e de gênero dentro do sintagma nominal – mesmo nos falares urbanos, é comum encontrar a marca de plural somente no determinante, como nas construções *As casinha amarela* e *Os menino saíram*.¹⁷
- (2) Variação na concordância verbo-nominal – que, em falares urbanos, atinge três pessoas do discurso (a 2ª pessoa do singular e do plural, e a 1ª pessoa do plural) e, em alguns falares rurais, chega a atingir a seis pessoas. Ex: *Você sabe* por *Tu sabes*/ *Vocês sabem* por *Vós sabeis* / *A gente vai* por *Nós vamos*.

¹⁷ Baxter, Lucchesi e Guimarães (1997) constataram que o estudo da mudança em tempo aparente aponta para um processo de implementação da regra de concordância de gênero no dialeto de Helvécia, comunidade afro-descendente isolada localizada no sul da Bahia. Na análise quantitativa dos dados as faixas etárias I (20- 40 anos), II (40-60 anos) e III (mais de 60 anos) alcançaram, no que concerne ao favorecimento das regras de concordância, pesos relativos de **.72**, **.56** e **.23**, respectivamente.

(3) Redução no uso do sujeito nulo – como consequência da redução da morfologia flexional do verbo.

(4) Estruturas de dupla negação – que são encontradas de duas formas: negação marcada no sujeito e no predicado, como na oração *Ninguém viu não*, em que o sujeito já possui carga semântica negativa; e duas partículas negativa na sentença, como em *Eu não vi ele não*. A primeira partícula possui duas realizações fonéticas (*não* e *num*), ao contrário da segunda.

(5) Uso de um mesmo verbo para expressar posse e existência – é o caso do verbo *ter*. Ex: Naquela praça *tem* duas árvores. (existência) / Eu *tenho* dois cachorros. (posse).¹⁸

(6) Focalização do constituinte através do deslocamento para o início da sentença – Ex: A *Maria*, eu não vi não. / O *livro*, ela acabou de ler ontem.

3.1.3 O sujeito pronominal

3.1.3.1 O parâmetro do sujeito nulo

O gerativismo, corrente teórica que busca descrever a Gramática Universal e definir quais são os princípios que a regem e quais as variações permitidas dentro dos mesmos, postula que o ser humano possui em seu aparato genético algo como a faculdade da linguagem, que é modular, ou seja, composta de módulos responsáveis por diferentes atividades. Esta teoria procura descrever as línguas e suas diferenças a partir de duas noções: *Princípios* – leis gerais válidas para todas as línguas naturais; e *Parâmetros* – propriedades que as línguas podem ou não exibir e que são responsáveis pelas diferenças entre elas (MIOTO, SILVA e VASCONSELOS, 2005, p.22-26).

Dentre os vários princípios existentes, há um princípio que diz que todas as sentenças finitas têm sujeito. Associado a esse princípio, há o **parâmetro do sujeito nulo**, o fato de o

¹⁸ Callou e Leite (2002), diante de um estudo feito em cinco capitais brasileira – Porto Alegre, Rio de Janeiro, São Paulo, Salvador e Recife, – concluem que o uso de *ter* por *haver* em construções existenciais já na década de 70 atinge até 89% entre os jovens. Em Salvador, entre adultos de 36 a 55 anos, o uso do *ter* existencial alcança 90% de frequência. Os resultados obtidos a partir de um *corpus* escrito colhido no ano de 1999 mostram que a substituição de *haver* por *ter* já chegou à língua escrita.

sujeito poder ou não estar sintaticamente presente, ainda que foneticamente vazio. A teoria **pro-drop** (teoria da queda do pronome), postulada por Chomsky (1981), estipulava que uma categoria nula [+anafórica - pronominal], conhecida na teoria gerativa como *pro*, é aceita na posição de sujeito de uma sentença finita somente se a morfologia verbal da língua em questão for rica o suficiente [+AGR] para permitir a sua recuperação. Esta seria a diferença básica entre línguas pro-drop (como o italiano ou o espanhol) e línguas não pro-drop (como o inglês). Enquanto em línguas pro-drop encontramos sentenças como as representadas pelos exemplos (37) e (38) abaixo.

(37a) Yo hablo español.

(37b) Hablo español.

(38a) Io parlo italiano.

(38b) Parlo italiano.

Em línguas não pro-drop, as sentenças sem o sujeito referencial realizado são agramaticais – cf. exemplo (39b).

(39a) I speak english.

(39b) * Speak English.¹⁹

O sujeito realizado também é encontrado, em algumas línguas, nas construções impessoais e existenciais, em que o sujeito é chamado de expletivo. O fato de uma língua ser pro-drop ou não também influencia a aceitabilidade do sujeito expletivo nulo. No inglês, por exemplo, constata-se a presença de sujeitos expletivos foneticamente realizados – cf. exemplos (40a) e (40b) –, obrigatórios nesta língua, enquanto no português, este mecanismo não é necessário, segundo pode ser visto em (41a) e (41b).

(40a) **It** rains.

(40b) **There** is a lot of people in the square.

¹⁹ O símbolo * antecedendo uma oração é comumente empregado na literatura lingüística para sinalizar uma construção agramatical.

(41a) ___ Chove.

(41b) ___ Tem muita gente na praça.

A alegação de que o sujeito pronominal nulo não é aceito em línguas em que a morfologia verbal é pobre foi modificada com a pesquisa de Huang (1984, 1989 Apud Duarte, 1995) em que se argumenta que mesmo em línguas como o chinês, em que não há morfologia verbal, o *pro* é aceito. Porém é necessário enfatizar que ausência de morfologia verbal não é o mesmo que ausência de concordância na sintaxe. Em casos como este, as ocorrências de *pro* são determinadas pela existência de estrutura de tópico [+ *close antecedente*]. O tópico é uma operação lingüística que consiste em fazer de um constituinte da frase o tema de que o resto da frase será o comentário, como nos exemplos (42a) e (42b):

(42a) A Maria, (ela) saiu ontem.

(42b) Futebol, eu não gosto não.

Com relação ao fator pro-drop, segundo Baptista (1995)²⁰, existem três tipos de línguas:

- a) Línguas que têm genuinamente sujeito pro-drop - só aceitam sujeitos nulos referenciais (sem precisar de antecedente), como o italiano.

(43) (Io) parlo italiano.

(Eu) falo italiano.

- b) Línguas que só aceitam sujeitos nulos não-referenciais e sujeitos nulos expletivos (44a), e em que sujeito nulo referencial (44b) é agramatical. São línguas chamadas de “semi pro-drop”, como o alemão.

(44a) Er sagte, daß ___ ihm scheint, daß Hans den Hund getötet hat.

Ele disse que (a) ele parece que Hans o cachorro matado tinha.

“Ele disse que **ec** parecia a ele que Hans tinha matado o cachorro.”

²⁰ Os exemplos (43) – (45) foram retirados de Baptista (1995).

(44b) * Er sagte, daß ___ den Hund getötet hat.

Ele disse que o cachorro matado tinha.

“Ele disse que *ec* tinha matado o cachorro”

c) Línguas “topic-drop” – que aceitam sujeitos e objetos nulos a partir de informações que são recuperadas pelo contexto, como no chinês.

(45) Zhangsan shuo [(ta) lai le]. (HUANG, 1989:188)

Zhangsan disse *pro* vir ASP.

“Zhangsan disse que *pro* vem.”

3.1.3.2 O sujeito pronominal nas línguas crioulas

Dentre as principais características das línguas crioulas está ausência de morfologia verbal de pessoa e número e o uso obrigatório de pronomes realizados (com uma possível exceção nas orações expletivas) mesmo em crioulos baseados em línguas românicas, que possuem morfologia verbal rica. Nestes casos, os itens mais referenciais, são mais frequentemente preenchidos que os itens não-referenciais.

Diversos pesquisadores, contudo, têm provado que, no que concerne ao parâmetro do sujeito nulo, as propriedades das línguas crioulas vão mais além do preenchimento obrigatório do sujeito pronominal de referência definida. Essas estruturas, como poderá ver visto adiante, no item 3.1.3.3, são compartilhadas tanto por línguas crioulas quanto por línguas que sofreram contato, como o português do Brasil..

Syea (1993) afirma que, no crioulo da Maurîtânia, há sujeito nulo mesmo sem morfologia verbal rica [-AGR] e sem estrutura de tópico [-close antecedent]. Em sua pesquisa são analisados exemplos tanto de sujeitos de referência definida quanto de referência não-definida. Na Maurîtânia, o sujeito nulo ocorre tanto em sentenças que apresentam as partículas de TMA quanto nas que não apresentam este mecanismo. Além disso, a categoria vazia também ocorre em sentenças finitas subordinadas (46), em que o sujeito tem referência arbitrária e não-específica, e em sentenças expletivas (47) e existenciais (48):

(46) Pyer ti dir [ec van puasõ dã bazar]
 Peter TNS dizer vender peixe em mercado.
 “Peter disse **ec** vende peixe no mercado”

(47) **ec** Posib Pyer lakaz.
 Possível Peter casa.
 “**ec** É possível que Peter esteja em casa.”

(48) **ec** Ena en voler dã lakaz.
 Existir uma faca em casa
 “**ec** Há uma faca em casa.”

Lipski (1999), estudando o Chabacano, um crioulo de base espanhola localizado nas Filipinas e muito falado na cidade de Zamboanga, postula que embora este crioulo seja denominado uma língua não pro-drop e não tenha morfologia verbal, sujeitos nulos podem ser autorizados em determinadas circunstâncias. Lipski afirma ainda que estas propriedades não são derivadas nem dos idiomas românicos que contribuíram para a formação deste crioulo nem do Filipino. O exemplo (49) apresenta uma construção com sujeito expletivo nulo. No Chacabano, pronomes sujeito nulos em orações simples não são aceitos, porém, quando o pronome sujeito é omitido em um contexto bem definido, que permite a sua identificação, como no exemplo (50), a maioria do falantes nativos reconhece este sujeito e classifica a oração como aceitável.

(49) **ec** Ya tene/tiene hente na mundo.
 ASP existir gente em mundo
 “Tem gente no mundo.”

(50) El muher tyene myedo y **ec** ya pregunta kosa le kyere
 A mulher tem medo e ASP pergunta coisa ele quer
 “A mulher tem medo e pergunta o que ele quer.”

Baptista (1995) afirma que o caso do Crioulo Caboverdeano é particular e que ele não se encaixaria em nenhum dos três tipos de línguas já apresentados, pois, além de sujeitos nu-

los não referenciais, nesta língua também são aceitos sujeitos nulos referenciais. No Crioulo Caboverdeano, sujeitos expletivos são obrigatoriamente nulos e são encontrados nas orações impessoais, em construções existenciais, em orações com cópula e em estruturas com verbo de alçamento. Outra característica desta língua é a de que a 3ª pessoa do singular com referência definida é preferencialmente nula quando a ela é atribuído um predicado permanente – cf. exemplo (51) – mas é obrigatoriamente realizada quando a ela é atribuída uma característica temporária, identificada pela cópula *sta* – cf. exemplos (52a) e (52b).

(51) (El) e um bom omi.

(Ele) é um bom homem.

(52a) El sta duente.

Ele está doente.

(52b) * Sta duente.

Está doente.

Na próxima sessão, será levanta a questão do português do Brasil e, ainda, como o parâmetro do sujeito nulo afetaria o sujeito indeterminado.

3.1.3.3 O comportamento do sujeito pronominal no português do Brasil

No português do Brasil, sujeitos nulos referenciais são permitidos desde que possam ser recuperados pela morfologia verbal ou pelo contexto em que se inserem – cf. exemplo (53) –, já sujeitos nulos de referência arbitrária, representados pelo exemplo (54a), só são aceitos desde que o *pro* tenha o mesmo índice referencial que o seu antecedente ou desde que também possa ser recuperado no contexto, como exemplificado em (55).

(53) Eles pensaram que **pro** poderiam ficar aqui também.

(54a) Ele_i disse que **pro**_i viria.

(54b) * Ele_i disse que **pro**_j viria.

(55) P: Maria vem à festa?

R: João_i disse que **pro**_j viria.

Normalmente, atribui-se a queda de sujeito pronome a um paradigma verbal rico, no entanto observa-se no português do Brasil uma significativa simplificação do paradigma flexional do verbo e, conseqüentemente, a existência de variação nas regras de concordância verbal. Após o desaparecimento da segunda pessoa *vós* e a inserção no paradigma verbal da segunda pessoa indireta *você*, cuja flexão verbal é a da terceira pessoa do singular, as flexões verbais no português do Brasil passam de seis para quatro, pois o pronome *tu*, mesmo quando utilizado, não é mais acompanhado por verbo flexionado. Quanto à primeira pessoa do plural *nós*, esta sofre concorrência, mesmo na norma urbana culta, do sintagma *a gente*, cuja flexão verbal também é da terceira pessoa do singular. Em variedades populares, somente a primeira pessoa do singular não sofre variação; já em falares mais isolados, como os de algumas comunidades afro-descendentes do interior da Bahia, a variação na concordância verbal chega a atingir as seis pessoas do discurso, não havendo nem mesmo diferenças entre singular e plural. O seguinte quadro, retirado de Lucchesi (2001), resume a mudança ocorrida:

Quadro 2: Paradigmas pronominais e flexionais no português do Brasil

Pessoa / N°	Pronome	Paradigma 1	Paradigma 2	Paradigma 3	Paradigma 4
1ª singular	Eu	Am o	Am o	Am o	Ama / Am o
2ª singular	Tu	Ama s	-	-	Am a
	Você	Am a	Ama	Am a	Am a
3ª singular	Ele/Ela	Am a	Ama	Am a	Am a
1ª plural	Nós	Ama mos	Ama mos	Ama / Ama mo(s)	Ama / Ama mo(s)
	A gente	-	Ama	Ama / Ama mo(s)	Ama / Ama mo(s)
2ª plural	Vós	Ama is	-	-	-
	Vocês	Ama m	Ama m	Ama / Ama m	Ama / Ama m
3ª plural	Eles/Elas	Ama m	Ama m	Ama / Ama m	Ama / Ama m

Fonte: Lucchesi, 2001, p. 118.

O paradigma 1 representa a norma padrão, restrita à língua escrita; o paradigma 2 refere-se à norma culta, em que já se inserem as formas *você(s)* e *a gente*; o paradigma 3 representa as variedades populares do português do Brasil e o paradigma 4 corresponde às formas encontradas nas comunidades afro-descendentes isoladas, que, segundo Baxter e Lucchesi (1995), em estudo feito sobre a variação na concordância verbal com a 1ª pessoa do singular na comunidade de Helvécia, comunidade afro-descendente isolada do sul da Bahia, apontam para implementação no uso das regras de concordância verbal.

Quanto ao sujeito expletivo e às construções existenciais, no português do Brasil, o sujeito nulo é obrigatório em construções existenciais (56), em construções impessoais (57) e em construções com verbo de alçamento, como na sentença apresentada em (58). Os itens (59) e (60) exemplificam sentenças muito comuns no português do Brasil, que a seguir serão discutidas.

(56) **cv** Há muitas árvores nessa cidade.

(57) **cv** Chove muito aqui.

(58) **cv** Parece que vai chover hoje.

(59a) **Você** tem muitas árvores nessa cidade.

(59b) **A gente** vê muitas árvores nessa cidade.

(60a) Você, se quer economizar, você tem que pegar um ônibus.

(60b) A gente, se quer economizar, a gente tem que pegar um ônibus.

Os pronomes *você* e a *gente* contidos nas sentenças em (59a) e (59b) possuem duas interpretações: podem tratar-se de sujeitos de referência arbitrária, ou seja, um sujeitos indeterminados de caráter genérico, como nas sentenças (60a) e (60b); ou, segundo Duarte (2003), podem tratar-se de sujeitos expletivos realizados. Segundo a autora, o português falado do Brasil estaria “passando uma mudança que envolve o Parâmetro do Sujeito Nulo no sentido de se tornar uma língua de sujeitos plenos” (p. 123). Em um processo de mudança como este, os itens referenciais cederiam à mudança antes dos menos referenciais.

3.1.4 Hipóteses acerca da indeterminação do sujeito no português popular

O processo histórico pelo qual passou o português do Brasil, debatidos no Capítulo 1 desta dissertação, licenciam a classificação dessa língua como resultante de um processo leve de contato entre línguas.

Os estudos acerca da transmissão lingüística irregular permitem-nos constatar que, em situação de contato massivo entre falantes de línguas mutuamente ininteligíveis, há redução/simplificação da morfologia da língua alvo, e uma posterior recomposição dos itens afe-

tados. Estes processos são variáveis quanto à sua intensidade, pois estão intrinsecamente ligados a fatores sociais, tais como acesso dos falantes à língua alvo.

Como foi visto nas sessões anteriores, línguas resultantes de transmissão lingüística irregular apresentam simplificação do paradigma verbal e um conseqüente preenchimento do sujeito pronominal. Este preenchimento, segundo os estudos que contemplaram o tema (cf. DUARTE, 1995), afetaria tanto o sujeito de referência definida quanto o sujeito de referência arbitrária. Espera-se, a partir desses dados, constatar no português popular no interior do estado da Bahia os seguintes fatos a cerca da indeterminação do sujeito:

a) alta realização dos sujeitos pronominais, em função da forte redução do paradigma flexional dos verbos resultante da aquisição defectiva do português, já constatada através de outras pesquisas (Cf. LUCHESE, 2004);

b) a baixa ocorrência da partícula indeterminadora *se*, sobretudo nas comunidades afro-brasileiras, ocasionada pela redução da morfologia da língua;

c) a alta freqüência e versatilidade da forma menos marcada – a categoria vazia (\emptyset + V3PS), decorrente também da perda de morfologia;

d) a existência de um *continuum* lingüístico proveniente dos processos de transmissão irregular da língua. No que concerne ao sujeito indeterminado, o *continuum* seria atestado pelas freqüências de uso da forma \emptyset + V3PS nas localidades analisadas neste trabalho. A categoria vazia apresentaria maior funcionalidade nas comunidades afro-brasileiras do que nas demais localidades, por manterem os traços da aquisição defectiva da língua em função do seu isolamento.

3.2 METODOLOGIA DE TRABALHO

O Projeto Vertentes do Português Rural do Estado da Bahia, dentro do qual esta pesquisa foi desenvolvida, é coordenado pelo Prof. Dr. Dante Lucchesi no Departamento de Letras Vernáculas do Instituto de Letras da Universidade Federal da Bahia (UFBa) e tem como objetivo estudar a realidade atual dos falares rurais do Estado da Bahia com base no enquadramento teórico-metodológico da Sociolingüística Variacionista. Implementado a partir do

ano de 2002, o Projeto Vertentes tem se dedicado à constituição de um acervo da fala do Português Rural do Estado da Bahia, em meio digital, e à realização de análises lingüísticas das diversas variedades da língua falada no interior do Estado.

3.2.1 As comunidades estudadas

Para este estudo foram analisados inquéritos de quatro comunidades afro-brasileiras e do português popular de um município do interior do Estado. O contraste entre essas se faz tanto pelo contexto histórico de sua formação quanto por serem as comunidades afro-brasileiras marcadas etnicamente, enquanto as demais localidades estudadas não o são.

As comunidades afro-brasileiras são assim chamadas por serem constituídas, em sua grande maioria, por descendentes de escravos africanos, reunidas em torno de um antigo quilombo ou em terras doadas pelos seus senhores. As comunidades analisadas nesta pesquisa foram: as comunidades de Barra e Bananal, no município de Rio de Contas; Sapé, em Valença; Helvécia, no município de Nova Viçosa; e Cinzento, no município de Planalto. No município de Santo Antonio de Jesus, foi recolhida a outra amostra de fala aqui analisada, que reúne entrevistas com indivíduos com pouca ou nenhuma escolaridade da sede e da zona rural desse município. A seguir, serão apresentados um pouco da história e da situação atual destas localidades.

3.2.1.1 Breve histórico e caracterização das comunidades afro-brasileiras

3.2.1.1.1 A comunidade de Helvécia

A comunidade de Helvécia tem origem na antiga Colônia Leopoldina, uma colônia suíço-alemã estabelecida em 1818, após o Governo Central ter concedido a três colonos alemães sesmarias ao longo das margens do Rio Peruípe, doze léguas acima de Nova Viçosa, na comarca de Caravelas, no extremo sul da Província da Bahia. Juntamente com outros colonos suíços e franceses fundam um conjunto de plantações que, dependendo da mão-de-obra escrava para o cultivo do café, prospera até a abolição, em 1888. No início da década de 1850, a Colônia Leopoldina era responsável por quase 90% da produção de café da Província. Conforme o que atestou o médico da colônia, em 1858, a sua população era composta de 200 brancos, principalmente suíços e alemães e alguns franceses e brasileiros, e 2000 negros, na

maior parte nascidos na colônia (TÖELSNER *apud* FERREIRA, 1988, p. 22). Esta situação, segundo alguns autores (cf. BAXTER; LUCCHESI, 1995) era especial, pois os senhores, como estrangeiros, não tinham como língua materna o português.

Significativo também o fato de a comunidade de ex-escravos da Colônia Leopoldina ter se fixado na região da colônia, de forma relativamente isolada, sem que tivessem se dispersado tanto quanto outras populações de ex-escravos após 1888.

No final da década de 1880, entretanto, uma combinação de fatores conspira para a rápida decadência da colônia: o advento da abolição e a expansão da cultura do café para regiões mais férteis da Província, que já havia se tornado o maior centro cafeeiro da Bahia.²¹

Apesar de diversos colonos terem deixado a região, em grande parte retornando à Europa, muitos dos ex-cativos e seus descendentes permanecem nas terras vizinhas das antigas plantações, praticando uma cultura de subsistência. Em 1897 é inaugurada a estação da ferrovia Bahia-Minas, e a vila de Helvécia passa a viver em função desta.

Em 1996 a sua população totalizava 6.474 habitantes – 15.108 destes viviam na zona rural e apenas 1.366 constituíam a população da pequena vila de Helvécia. Dos seus 13.310 habitantes com mais de cinco anos, 7.203 eram alfabetizados; o que corresponde a uma taxa de alfabetização de cinquenta e quatro por cento, que prevalece tanto na zona rural (com 6.484 indivíduos alfabetizados, em um total de 11.972), quanto na zona urbana (719 indivíduos, em um total de 1.338). Na faixa da população com mais de sessenta anos, essa taxa de alfabetização cai para vinte e dois por cento - 195 em um total de 887.²²

Atualmente, pouco há da antiga presença dos antigos colonizadores europeus, mas a grande concentração de escravos que remonta as origens da antiga Colônia Leopoldina ainda se faz notar na composição étnica da população local.

A comunidade de Helvécia, atualmente, está inserida normalmente na economia da região do extremo sul do Estado da Bahia. Sua economia é baseada fundamentalmente na agricultura, na pecuária e na indústria da celulose. O turismo balneário tem se desenvolvido, sobretudo na sede do município de Nova Viçosa, situada na costa.

A maioria dos habitantes da comunidade entrevistados dedica-se à cultura de subsistência e trabalha no fabrico artesanal da farinha de mandioca em pequenas propriedades agrícolas familiares, vendendo todo o eventual excedente nas feiras de Helvécia ou de Teixeira de Freitas. Raros são os que empregam alguma maquinaria e adubos e fertilizantes químicos

²¹ Segundo dados do Governo do Estado da Bahia (1978). *A inserção da Bahia na evolução nacional - 1ª etapa: 1850-1889 - atividades produtivas*. Salvador: Secretaria do Planejamento, Ciência e Tecnologia.

²² Fonte: IBGE, Contagem da população 1996.

no cultivo de horti-fruti-granjeiros. Há também, na região, o cultivo de feijão, milho, arroz, abóbora, batata, e outros legumes.

A implantação de grandes plantações de eucalipto, ramo no qual se empregam eventualmente alguns dos moradores mais jovens da localidade, entretanto, alterou a fisionomia da região e desalojou muitas famílias que viviam do cultivo em suas pequenas propriedades. Esses indivíduos e seus descendentes, agora, trabalham em terras próximas de parentes ou se empregaram em grandes propriedades agrícolas. Muitos buscam o mercado de trabalho urbano, como vigilantes, operários da construção civil, empregadas domésticas, por exemplo, nos centros urbanos regionais, como Teixeira de Freitas-Ba e Nanuque-MG, e nos grandes centros do sul do país.

3.2.1.1.2 A comunidade de Rio de Contas

O termo Rio de Contas é utilizado, dentro da metodologia do Projeto Vertentes, para denominar as comunidades de Barra e Bananal, que se situam no município de Rio de Contas, no estado da Bahia, e se distanciam por apenas dois quilômetros.

A região começa a ser ocupada na última década do século XVII, através do estabelecimento de uma rota de viagem entre Goiás e o Norte de Minas e a cidade de Salvador, capital da então Província da Bahia. Um pequeno povoado, de nome *Creoulos*, situado em um planalto da serra das Almas, na margem esquerda do Rio de Contas Pequeno, atual Rio Brumado, é estabelecido como ponto de pouso nesta rota. O desenvolvimento da mineração, propiciado pela descoberta de veios e cascalhos auríferos, no Rio de Contas e em seus afluentes e serras circunvizinhas, aliado ao aumento da população do povoado culminam na criação, em 1718, da primeira freguesia do Alto Sertão Baiano - ou Sertão de Cima -, denominada Santo Antônio de Mato Grosso.

Em 1722, o Conselho Ultramarino cria a Vila de Nossa Senhora do Livramento das Minas de Rio de Contas, onde os jesuítas haviam erigido uma igreja em devoção a Nossa Senhora do Livramento - doze quilômetros abaixo do antigo povoado de Creoulos. Aí, atualmente, se situa a cidade de Livramento do Brumado. Porém, em 1745, a mudança da sede da vila para o antigo povoado de Creoulos é autorizada por uma Provisão Régia e a vila passa a se chamar Vila Nova de Nossa Senhora do Livramento e Minas do Rio das Contas, enquanto que a antiga sede passou a ser conhecida por Vila Velha. A mesma Provisão também eleva a nova vila à categoria de freguesia e transfere para aí a sede da freguesia de Santo Antônio

de Mato Grosso, com a denominação de freguesia do Santíssimo Sacramento das Minas do Rio de Contas. O Rio de Contas, portanto, recebia todos os recursos e benefícios da atividade mineradora, o que se refletia no seu crescimento urbano.

A comunidade de Bananal teria sido fundada por volta do século XVII por negros, que escolheram as cabeceiras do rio Brumado para viver após o naufrágio de um navio negreiro vindo da África. Estes foram, depois de já estabelecidos no local, escravizados por bandeirantes, chefiados por Antônio Raposo Tavares. Hábitos já desenvolvidos, como a prática da cultura de subsistência e o cultivo de suas tradições, foram substituídos pelo trabalho escravo na mineração.

Em 1840, o nome do município muda para Minas do Rio de Contas; e, em 1931, o Município passou a se chamar, simplesmente, Rio de Contas. O seu crescimento vai se estagnando com o esgotamento progressivo dos veios e cascalhos auríferos.

Barra e Bananal são povoados afro-descendentes, e seus habitantes, assim como os seus antepassados, praticam ainda a agricultura de subsistência. A endogamia é uma prática comum nos dois povoados.

Atualmente, o isolamento das duas comunidades é quebrado pela indústria do turismo instalada em torno da região do município de Rio de Contas, porém seus habitantes ainda vivem em condições precárias de saneamento e educação. Algumas das tradições de origem africana se perderam por causa do contato com a cultura branca européia e, especialmente, por causa dos valores do catolicismo, que se tornou a religião predominante.

3.2.1.1.3 A comunidade de Sapé

A comunidade chamada Sapé recebe dos seus habitantes duas denominações: os mais antigos chamam-na de *Sapé Grande*, enquanto os mais jovens de *Sapé Alto*. A comunidade localiza-se no Distrito do Município de Valença, na região do Recôncavo Baiano, e tem seu acesso por uma estrada de terra cujo percurso de 25 quilômetros inicia-se em direção inversa à que leva à cidade de Valença, no entroncamento da BR-101.

Segundo alguns moradores da cidade, a comunidade formou-se algum tempo após a assinatura da abolição dos escravos. O Sr. João Barreto, morador da região, conta que, em 1800, as terras que compreendiam Sapé, Rapa Tição e Tabuado pertenciam ao Sr. Miguel Elia. Com a sua morte, se deu a divisão das terras entre seus dois filhos, que as venderam. O que se sabe sobre os costumes dos habitantes nesta época faz parte de histórias contadas pe-

los antepassados dos atuais moradores e passadas para seus parentes. O Sr. João Barreto diz que, segundo o que contava seu avô, o primeiro a chegar à região após a família de Miguel Elia, as atividades econômicas da época restringiam-se ao cultivo da mandioca e à criação de gado – grande parte de suas terras eram reservadas para pasto. A família do Sr. João Barreto guarda relíquias desta época, em que, segundo o mesmo, existiam, nos últimos anos do século XIX, senzala e troncos onde os escravos eram castigados na fazenda da qual seus familiares tomaram posse. Tratam-se de algemas e correntes que lá foram encontradas.

Na época, a região do Sapé teve que ser desmatada e não havia qualquer resquício de plantação. Foi encontrado na terra apenas sapé, uma espécie de capim da família das gramíneas, o qual utilizaram para cobrir suas moradias.

Após a abolição, muitos negros ex-escravos permaneceram trabalhando nas fazendas. Posteriormente, como os fazendeiros da região não gozavam de uma boa situação econômica, para quitar suas dívidas com os ex-escravos, "doaram"-lhes pedaços de terra improdutivos, mais altos, sem trato e distantes da água. Esses permaneceram no local, constituíram família casando entre si mesmos e aumentaram a população local. Mantiveram-se, entretanto, isolados, como o que aconteceu com as comunidades de Helvécia e Rio de Contas.

A comunidade conta atualmente com, aproximadamente, 100 habitantes, na região mais concentrada, e é constituída, predominantemente, como em todas as sociedades mais recentes, por jovens e crianças. A energia elétrica chegou à região há apenas 10 anos, contudo a maioria das famílias não pode utilizá-la, pois se restringe à rua principal. Mas os aparelhos de TV são bastante encontrados, ligados, às vezes, à bateria de carro em lugares onde a rede elétrica não chegou ainda. A comunidade não é atendida ainda pelos serviços de saneamento básico -, quem tem banheiro em casa é considerado rico. Homens, mulheres e crianças trabalham na roça ganhando diariamente entre R\$ 4,00 e R\$10,00 em época de plantio ou colheita. A agricultura de subsistência também é praticada por algumas famílias, que cultivam feijão e mandioca, utilizada na fabricação da farinha, cujo excedente vendem na BR 101. Há também alguns pés de cravo e cacau plantados.

Não há, na população local, a ambição de sair da comunidade em busca de melhores condições de vida, como se pode constatar nas comunidades de Helvécia e Rio de Contas.

A religião que predomina na comunidade é a Católica. Apesar das suas origens, os informantes, durante as entrevistas, afirmam com orgulho não ter entre eles qualquer seguidor de cultos africanos.

3.2.1.1.4 A comunidade de Cinzento

A comunidade de Cinzento, considerada uma comunidade afro-brasileira, tem origem num quilombo incrustado na região sudoeste da Bahia, próximo da cidade de Planalto, a 77 km de Vitória da Conquista e 450 km da cidade do Salvador. Segundo relatos dos antigos moradores, os fundadores de Cinzento são da região da Chapada Diamantina, mais precisamente do antigo "Arraial dos Crioulos" e o sobrenome *Pereira Nunes*, predominante em Cinzento, é o mesmo de um antigo proprietário de escravos da região de Rio de Contas. A moradora Grazielle de Lourdes Novato Ferreira afirma que a vinda dos primeiros negros para o Cinzento está associada às comunidades estabelecidas à margem do Rio Gavião por volta de 1810 e 1860. Ana Isidora (107 anos) sugere que os primeiros moradores vieram de forma clandestina, pois a vinda se deu "à meia-noite, terça-noite". Os mais velhos afirmam que seus pais contavam ter a obrigação de se ajoelharem "em frente dos donos". Ana Isidora vai mais além e conta que sua bisavó era "caboca do mato e minha avó Maria foi pegada no mato"

Lourenço Pereira Nunes desmatou as primeiras terras e tomou posse delas, segundo o senhor Saviano Pereira Nunes, nascido e criado no Cinzento. Tercílio, também morador da comunidade, afirma que os primeiros desbravadores foram o já citado Lourenço Pereira Nunes e Sérgio Pinheiro dos Santos. O nome Cinzento, conta Ana Isidora, provém de um boi gordo, "de tão gordo parecia cinza", que apareceu milagrosamente para alimentar os primeiros moradores.

A endogamia, como em demais localidades isoladas, ainda é hábito dos moradores e é uma forma de manter as terras na família em caso de divisão.

3.2.1.2 Breve descrição da cidade de Santo Antonio de Jesus

Santo Antônio de Jesus, um dos mais importantes municípios do Recôncavo Baiano, se localiza a 187 km de Salvador e faz limite com Aratuípe, Conceição do Almeida, Dom Macedo Costa, Elízio Medrado, Laje, Nazaré, São Felipe, Muniz Ferreira e São Miguel das Matas e Varzedo. Quanto à distribuição da sua população, segundo dados do censo demográfico de 2000, dos aproximadamente 73.400 habitantes do município, cerca de 66.250 residiam na zona urbana e 11.150 na zona rural. Em 2007, segundo os dados do IBGE, o município conta com uma população de 84.256 habitantes e uma área de 259 m².

O município originou-se durante os séculos XVII e XVIII através do desenvolvimento da freguesia da Paróquia de Santo Antônio em torno da capela, onde atualmente encontra-se a praça Padre Mateus e a Matriz de Santo Antônio de Jesus. Situada numa área de terras férteis e de grandes recursos pluviais, logo foram estabelecidas plantações de mandioca e engenhos de cana-de-açúcar ali.

A freguesia Paróquia de Santo Antônio foi elevada à categoria de Vila em 29 de maio de 1880 e, posteriormente, no mesmo ano em que ocorreu a inauguração da estrada de ferro da cidade, mais precisamente em 30 de julho de 1891, a Vila é elevada à categoria de Cidade. Em 1875, segundo a Enciclopédia dos Municípios Brasileiros (1958, p.36), a população de Santo Antônio contava com 9.654 habitantes, em sua grande maioria negra, 4.000 escravos, e 3.000 estrangeiros residentes em seu território. O número total de analfabetos era de cerca de 8.320 pessoas.

A história do município é marcada tanto por embates políticos – pela existência de centros abolicionistas, republicanos, conservadores e liberais – quanto pela riqueza das suas tradições religiosas e culturais. Dentre estas manifestações, têm destaque os festejos do padroeiro do município, em que é realizada a famosa trezena de Santo Antônio, os dias de São João, São José e São Benedito.

O município de Santo Antônio de Jesus vive basicamente, em termos econômicos, do comércio local, que vem se consolidando como um grande pólo comercial, para onde se dirige um grande número de consumidores atraídos pela grande oferta do ramo de confecção, móveis e eletrodomésticos. No ano de 2003, foi inaugurada uma grande indústria do ramo de calçados, porém sempre houve micro-indústrias, voltadas para a produção de fogos de artifício e de vassouras na região.

Uma das grandes atrações da localidade é a grande feira-livre que ocorre no centro da cidade. Além de atrair um grande número de pessoas que vai em busca de produtos como roupas e utensílios domésticos vendidos por ambulantes, a feira também é conhecida por fornecer produtos cultivados na zona rural do município. Desta forma, muitos habitantes da zona rural de Santo Antonio de Jesus deslocam-se com frequência para a zona urbana do município para obter seu sustento através da venda de produtos agrícolas. Muitos habitantes da zona rural, entretanto, têm migrado para a zona urbana do município buscando melhores condições de vida e ofertas de emprego. Apenas um pequeno número de habitantes ainda sobrevive no campo graças à agricultura de subsistência ao cultivo da laranja e de limão e à venda de suas mudas na sede do município e em outras cidades.

3.2.2 O corpus

O *corpus* utilizado nesta pesquisa é composto pelo *Corpus Base do Português Afro-brasileiro do Estado da Bahia* e por 24 entrevistas coletadas na cidade de Santo Antônio de Jesus que pertencem ao *Acervo de Fala Vernácula do Português Rural do Estado da Bahia*.

O *Corpus Base do Português Afro-Brasileiro do Estado da Bahia* é constituído por 24 entrevistas do *Acervo de Fala Vernácula do Português Afro-Brasileiro do Estado da Bahia*. Deste foram selecionados dois informantes de cada uma das três faixas etárias, um de cada sexo, de cada uma das quatro comunidades que integram o Acervo, de modo que a diferença geracional na amostra pudesse ser equilibrada.

Ao Acervo de Fala Vernácula do Português Rural do Estado da Bahia pertencem as entrevistas coletadas em dois Municípios: Poções e Santo Antônio de Jesus. Foram realizadas em cada um dos municípios 24 entrevistas – 12 na sede do município e 12 na zona rural. Para esta pesquisa, entretanto, não se tornou necessária a utilização das entrevistas realizadas no município de Poções, visto o número satisfatório de ocorrências encontradas.

Ambos os acervos foram constituídos dentro do Projeto Vertentes e segundo a metodologia do mesmo, que pode ser consultada na página do Projeto Vertentes (www.vertentes.ufba.br).

3.2.2.1 As entrevistas

Todos os inquiridos, captados entre os anos de 1992 e 2002, têm duração de 40 a 60 minutos cada. Independente do meio original em que foram gravadas²³, as entrevistas passaram por processos de digitalização para que a qualidade do áudio fosse melhorada²⁴. A transcrição das entrevistas foi realizada conforme uma chave de transcrição específica, elaborada segundo as necessidades e os objetivos do projeto.²⁵ Para a constituição do acervo de fala do Projeto Vertentes, foi feito um recorte sincrônico, e a análise se dá por estudo em *tempo aparente*, através do qual se procura observar a mudança em função da faixa etária dos informantes.

²³ As entrevistas das comunidades de Rio de Contas, Helvécia e Cinzento foram gravadas em fitas cassete e depois digitalizadas. As entrevistas de Sapé e Santo Antônio foram feitas diretamente em meio digital.

²⁴ Na página do Projeto Vertentes, encontram-se trechos de áudio das entrevistas utilizadas pelo projeto bem como a transcrição dos mesmos em www.vertentes.ufba.br/audio

²⁵ A chave de transcrição consta na página do Projeto Vertentes na internet, no endereço www.vertentes.ufba.br/chave.

3.2.2.2 Os informantes

Os informantes são distribuídos por seis células, resultantes do cruzamento de duas variáveis sociais, *sexo* e *faixa etária do informante*, segundo a metodologia da sociolinguística variacionista quantitativa. Além destas duas variáveis já citadas, foram controladas também outras duas variáveis sociais, embora não tenham servido à elaboração das células: *escolaridade do informante* e *estada fora da comunidade por seis meses ou mais*. Todos os informantes, escolhidos aleatoriamente entre os habitantes das comunidades, são nascidos no local ou foram para lá levados em seus primeiros anos de vida e possuem pouca ou nenhuma escolaridade.

3.2.3 O suporte quantitativo

Um dos pressupostos básicos da sociolinguística é o de que a variação linguística não é aleatória. É fundamental, portanto, responder ao **problema das restrições**, identificando “conjuntos de circunstâncias linguísticas e sociais [...] que tendem a favorecer ou desfavorecer o uso de uma ou de outra variante, assim se correlacionando ao uso da variável dependente sob análise” (NARO, 2003, p. 148).

Modelos matemáticos – dentre eles o modelo aditivo o modelo multiplicativo e o modelo logístico ou misto – foram desenvolvidos com o intuito de tratar estatisticamente dados linguísticos (Cf. NARO, 2003), contudo o que apresenta uma síntese de dados dos anteriores é o mais utilizado atualmente.

A partir da codificação das ocorrências da variável dependente em função das variáveis independentes, realiza-se a análise quantitativa dos dados através do pacote de programas VARBRUL – do inglês *variable rules* – (Cf. SANKOFF, 1988; PINTZUK, 1988). Primeiramente são fornecidas as frequências relativas ao uso de cada uma das variantes que compõem as variáveis independentes. Como produto final tem-se valores denominados *pesos relativos*, que medem a influência de cada fator das variáveis independentes sobre a variável analisada. Fatores que atingem pesos relativos acima de 0.50 favorecem a variável dependente, enquanto os pesos abaixo de 0.50 desfavorecem.

Contudo, os valores obtidos com este programa são apenas estatísticos; o linguísta deve, portanto, atribuir o valor linguístico dos dados e interpretá-los.

3.2.4 Levantamento das ocorrências

Para o propósito desta pesquisa, foram levantadas as ocorrências de sete estratégias de indeterminação: os pronomes lexicais *a gente*, *nós*, *você* e *eles* e as formas verbais da 3ª pessoa do plural da 1ª pessoa do plural sem sujeito realizado foneticamente e sem uma ligação a uma referência específica anterior (doravante, $\emptyset + V3PP$ e $\emptyset + V1PP$, respectivamente). Nesse mesmo contexto, foram depreendidas as formas verbais da 3ª pessoa do singular (doravante $\emptyset + V3PS$). A nomenclatura categoria vazia refere-se ao sujeito nulo da construção $\emptyset + V3PS$. As referidas variantes são exemplificadas a seguir.

a) gente

(61) *A gente* puxava a roda, ia lá, bulia com o [burinete], celava a mandioca, fazia a farinha.

b) nós

(62) *Aí nós* ‘panhava peda que aqui tinha um home que chamava Bernadino Lauro, que morreu, e tinha a peda dela, sabe?’

c) você

(63) Hoje só tem convessa fiada. *Você* foi um forrozinho, ora... com pôco, turma tá discutino.

d) eles

(64) Eu não me engano não, mas acho que aqui, eu acho que é um tal de... *qu'eles* fazem um arquerão...

e) $\emptyset + V3PP$

(65) Me *dissero* que *vão* pegá de noite... que *vão* dá um côro nele de noite.

f) $\emptyset + V1PP$

(66) Então, *quando* tem que fazê nossa vigília... Né, *aproveitamos* fazendo vigília, né? Fazendo retiro naquele dia por que ININT. É um período de férias.

g) Ø + V3PS

(67) É... que *diz* que de noite não *pode* tomá muito tempo pa intervalo não. Poque *saía* dez hora.

Ocorreu também, porém em número reduzido, a utilização do pronome de tratamento *o senhor* para expressar indeterminação, bem como dos pronomes *eu e ele*, porém, por sua baixa freqüência, as ocorrências dessas variantes foram descartadas na análise quantitativa.

- o(a) senhor(a)

(68) A *senhora* ININT a *senhora* ganha quinhento reai... a *senhora* vai tê direito... a *senhora* vai tê... a *senhora* vai tê a facilidade de tirá do seu dinheiro trezento pa pagá, num tem.

- eu

(69) Gente de tempo é assim.... Se *eu* dô o senhô má criação, eles me dão um côro chateado...

- ele

(70) Já passô medino aí, num é hoje... isso já tá com mais de quatro mês *qu'ele* tá medino isso, e até hoje não aparece, né?

Na constituição dessa base de dados, não foram consideradas ocorrências em que:

a) o pronome tinha como antecedente um item lexical;

(71) A Schincariol também faz a mesma coisa, a gente pega o produto deles pra vendê, e *eles* dão um cantô pra gente...

b) a ocorrência tinha como referência um sintagma como “o pessoal”, “a pessoa”, “o cara”, mesmo quando esta se encontrava na fala do documentador;

(72) DOC1: E a *pessoa* sente mesmo o calô, num sente?

INF: Sente...

- c) casos de interpretação ambígua ou que não forneciam dados suficientes para que fosse confirmado o sujeito;

(73) E a gente fez... o *peessoal* fez a... a... uma parte, né?

- d) sujeitos em que há referências locativas, tais como nas expressões “eles de lá” e “o povo de lá”;

(74) O Paramirim...o do arraiá...o do arraiá... fica... é... é bem pra lá. O ôto é...o ôto é...o ôto é longe, mas é acolá mermo. É um burrado aí nessa terra, moço, mas *eles de lá*... eles de lá já ‘costumaro... [Eles] já ‘costumaro pra lá.²⁶

A partícula *se* teve tratamento qualitativo (ver item 4.4), pois o seu apagamento seria, segundo a metodologia empregada nesta pesquisa, um indicador da perda de morfologia gramatical que afetou os falares populares através do processo de contato e do aprendizado deficiente da língua portuguesa, o que, se constatado, contribuiria com as hipóteses acerca do vernáculo das comunidades rurais defendidas pelo *Projeto Vertentes*.

A chave de codificação utilizada nesta pesquisa foi fruto de um estudo piloto sobre a indeterminação do sujeito pronominal no português urbano culto, realizado no ano de 2004. A variável dependente foi codificada segundo oito variáveis lingüísticas e cinco variáveis sociais, apresentadas a seguir.

3.2.5 Variáveis lingüísticas

Entre os contextos lingüísticos que poderiam estar afetando a escolha da estratégia para expressar a indeterminação do sujeito, incluímos nesta análise os seguintes: (i) realização fonética ou não do sujeito indeterminado; (ii) desinência de pessoa e número presente ou não na forma verbal em foco; (iii) tipo de verbo; (iv) tipo de frase; (v) nível de referencialidade do agente; (vi) modo verbal; (vii) inclusão do falante no universo de referência do sujeito indeterminado; (viii) paralelismo formal. Nas seções seguintes, faremos um breve comentário sobre cada uma dessas variáveis lingüísticas explanatórias.

²⁶ Neste exemplo, o item *eles de lá* refere-se diretamente às pessoas de Paramirim. Trata-se, portanto, de um sujeito determinado.

3.2.5.1 Realização fonética do sujeito indeterminado

Esta variável não se aplica a estratégias verbais como \emptyset + verbo em 3ª p.p. ou \emptyset + verbo em 1ª p.p., por exemplo, e quando uma ocorrência de sujeito nulo é imediatamente antecedida por um pronome que assume função indeterminadora – cf. exemplo (33), é tomada como ocorrência desse. Desta forma, é possível, de modo confiável, distinguir as ocorrências de pronomes nulos das ocorrências de categoria vazia, apesar da ausência de flexão verbal.

Esta variável foi proposta pelo seu caráter descritivo e por fornecer dados relevantes a pesquisas acerca do princípio “Evite pronome” (Cf. DUARTE, 1995), que, ao que se sabe, afeta também a realização do sujeito de referência arbitrária.

As possibilidades de realização desta variável são:

a) Sujeito realizado

(75) *É, a gente procura onde tá a terra mais molhada, a gente sabe que dá água mais rápido, né?*

b) Sujeito não-realizado

(76) *Aí, eles fizeram logo uma simpatia, \emptyset partiu a cabeça da cobra e tal, \emptyset tirô o mio-lo... \emptyset amarrô em cima...*

3.2.5.2 Desinência verbal

O paradigma flexional dos verbos no português do Brasil tem sofrido uma redução significativa provocada pela substituição dos pronomes *tu* e *vós* pelas formas *vocês* e *vocês*. Essa erosão tem se agravado com a implementação da expressão *a gente*, o que faz com que restem em nosso paradigma flexional apenas três formas distintas (DUARTE, 1995, p.18).

Desta forma, esta a variável *desinência verbal* foi proposta por permitir mensurar a relação entre a redução do paradigma flexional do verbo e a escolha do falante por formas menos marcadas. A princípio, abrangia também a desinência de 1ª p.s., porém, pela pequena ocorrência do pronome *eu* em função indeterminadora e, conseqüentemente, por não podermos analisar os dados concernentes ao uso deste pronome, esta variante foi excluída da análise dos dados.

a) desinência zero

(77) Nós *apanhava* \emptyset que num *deveria* \emptyset fazê isso com o mais velho.

b) desinência da 1ª pessoa do plural

(78) Nós *temos* o período de... de... de retiro, né?

c) desinência da 3ª pessoa do plural

(79) Eles não *prestam* atenção, não *dão* ‘portunidade pa aquelas pessoas.

3.2.5.3 Tipo de Verbo

Nesta análise, adotou-se a classificação apresentada em Mira Matheus *et alia* (2001), com algumas adaptações. Dessa forma, foram definidos quatorze tipos de verbo:

a) transitivo *discendi*

(80) ...ninguém qué dá um dia de trabalho. *Diz* logo que é ladrão...

b) transitivo cognitivo

(81) Quando a gente tá jovem, num *pensa* nas consequência.

c) transitivo de ação pontual

(82) A gente comprô esse comprometido.

d) transitivo de ação não-pontual

(83) E mercado a gente *aprende* um muitas coisa, né?

e) intransitivo

(84) 'Que aqui a gente *trabalha* muito.

f) intransitivo inacusativos

(85) Não, a gente vai mah tem sodade daqui, *nasceu* aqui... tem sodade do lugá da gente.

g) verbo de ligação/estativo

(86) Quando a gente *tá* assim, a gente *qué* tê uma esperança.

h) verbo de movimento

(87) Deu cinco hora, a gente *veio*.

i) verbo locativo

(88) Tem algum lugá aí que há bagunça, mas, principalmente, essa rua aqui que nós *moramos*.

j) verbo de posse

(89) Se *tivé* um real, vem de moto, se num *tivé*, vem andano mesmo.

l) verbo suporte/leve

(90) Aí, pronto: palestrava, orava, começava... conversava, *dava* conselho...

m) verbo causativo

(91) Aí me *mandaro* que eu fosso pa Salvadô urgente.

n) verbo existencial

(92) Bom, aqui a gente *tem* um grupo de... que trata o grupo de mensagêro da paz.

o) verbo sem carga semântica/marcador conversacional

(93) A gente ranca, descasca, lava, aí *vai* lá, joga no motô...

3.2.5.4 Tipo de frase

Esta variável também não se mostrou produtiva, pois, em função da natureza das entrevistas (realizadas de modo a favorecer e incentivar somente o discurso do falante, e não um diálogo), poucas foram as ocorrências de frases interrogativas – quase todas elas foram realizadas a partir do turno do documentador – cf. exemplo (53). Os tipos de frases adotados são os tipos já consagrados na literatura lingüística:

a) declarativa

(94) Vendeu laranja, ‘bacaxi, no máquina de ferro, de Maria Fumaça.

b) negativa

(95) A gente num se arrisca não.

c) interrogativa

(96) DOC: E a senhora conhece aqui com'ê que formô, aqui, o arraial?...tem...

INF: Formô aqui?

3.2.5.5 Nível de referencialidade do agente

Já tratada por alguns pesquisadores como grau de indeterminação (Cf. CUNHA, 1993), esta variável distribui-se por quatro níveis: *genérico*, *parcialmente definido*, *específico indefinido*, e *eu ampliado* (ver Capítulo 2, item 2.3).

Segundo a metodologia aqui adotada (Cf. PONTE, 2006), ao nível *genérico* cabem as construções indeterminadoras nas quais as ações verbais podem ser desempenhadas por qualquer pessoa. Ao *grupo parcialmente definido* pertencem construções em que se pode deprender, através do contexto, uma referência; esta, no entanto, não é explicitada e não permite uma delimitação clara. O nível *eu ampliado* é representado por construções em que figuram como sujeito a primeira pessoa dilatada – nelas o falante utiliza-se de formas de primeira pessoa do plural para referir-se a si mesmo. Finalmente, o nível *agente específico indefinido* trata das construções clássicas em que, de fato, o sujeito da ação é desconhecido ou em que não há interesse no seu conhecimento. Esses quatro níveis são, a seguir, exemplificados.

a) genérico

(97) *Você* não pode fazê muito filho mode disso. Tem despesa, *você* fez fora de... filho na rua, não pode!

b) grupo parcialmente definido

(98) Fazendo retiro naquele dia por que ININT. É um período de férias. Aí, sempre *nós* fazemos, mas *nós* não comemoramos o natal.

c) eu ampliado

(99) *A gente* chegava fazia calo aqui nos dedo, aí. Nesses calo eu tinha aqui nunca demanchô do cabo do rôdo que...

d) agente específico indefinido

(100) Foi um pobrema que teve no cólo do útero, um sisto. Aí *descobriro* e aí me *mandaro* que eu fosso pa Salvadô urgente.

3.2.5.6 Modo verbal

Esta variável trata do modo que distingue a referência a fatos e eventos reais da referência, denominado como modo *realis*, a fatos e eventos possíveis, hipotéticos e irreais, denominado como modo *irrealis*. Essa distinção modal pode ser exemplificada da seguinte maneira:

a) *realis*

(101) *A gente* prantô ali. Tem um pedaço ali prantado, mas não vai prestá não.

b) *irrealis*

(102) *A laranja*, *você* limpô, Ø pode largá lá e... que aquando chegá a... o tempo de adubá, Ø adubô, pronto.

3.2.5.7 Inclusão do falante no universo de referência do sujeito indeterminado

Esta variável foi estruturada considerando a possibilidade de o sujeito indeterminado incluir ou não o falante, como se pode ver nos exemplos abaixo.

a) falante incluso

(103) DOC: E as mulheres aqui usam o pano na cabeça é por esse motivo ou por ôtro motivo?

INF: Assim, pá... *A gente* tá limpano assim... trabaiano na roça... *A gente* sempre usa o cabelo mais de fora assim é quando *a gente* vai, assim, viajá, né?²⁷

²⁷ A informante é uma mulher e, durante a sua fala, inclui-se no grupo das mulheres que usam lenço na cabeça.

b) falante não-incluso

(104) É, essa aqui é eu, poque eu tava sentada aqui e Ø *bateu* assim, a máquina bem pertin. Eu sentada aqui e ... e quem bateu o foto tava sentado aqui desse lado, né.

3.2.5.8 Forma antecedente

Esta variável concerne à relação de interdependência apresentada por variáveis dependentes no discurso. “O paralelismo formal tem se mostrado pertinente para um número suficientemente grande de línguas naturais [...] e para um número também suficiente grande de fenômenos do português falado no Brasil [...]”(NARO e SCHERRE, 1993, p. 11)

A menção anterior aqui só foi considerada quando constatado o sujeito indeterminado da oração imediatamente anterior em um mesmo contexto e em um mesmo turno. Os demais casos, como as ocorrências que iniciaram uma seqüência ou quando houve mudança do referente, foram classificados como primeira menção.

Como já citado anteriormente, a baixa ocorrência das formas *eu*, *ele* e *senhor* impossibilitou que fossem encontrados casos de sujeito indeterminado antecidos pelas mesmas.

a) primeira menção

(105) Por que, acho assim, que, mesmo assim a pessoa sendo... não tem aquela... a-quele nível de série, aquela coisa. Acho que Ø *deveria* dá oportunidade.

b) com *a gente* explícito na oração anterior

(106) Pensa, quano vai *a gente* pensa... pensa ni ladrão, Ø **pensa** ni onça, de tуди que tem dento de casa.

c) com *a gente* implícito na oração anterior

(107) Aí amanhã a gente assistia o...o procissão, Ø se num quisesse Ø **ia embora** mais cedo...

d) com *nós* explícito na oração anterior

(108) Quando ‘doce, *nós* leva pá hospitale, Ø **cuida** lá no hospitale.

e) com *nós* implícito na oração anterior

(109) É, o gravatazim, nós tirava, \emptyset *batia*, \emptyset **trançava** cortinha pa vendê a quinhento réis.

f) com *você* explícito na oração anterior

(110) Tano seco, cê proveita, que nem tá o soli aberto aí. *Ocê* proveita e \emptyset **vai ranca-**
no e sacudino assim pá enxugá.

g) com *você* implícito na oração anterior

(111) Você acende sua foguera na sua porta, praticamente ninguém vai na sua casa, cê tem que ir diretamente...ou fe... \emptyset *fecha* sua casa e \emptyset **sai**.

h) com *eles* explícito na oração anterior

(112) Que *eles* pega, **eles cata** as bainya...

i) com *eles* implícito na oração anterior

(113) Aí eles da aí assim minha senhora, \emptyset *dá* os meterial e aí \emptyset **dá** as pessoa...

j) com verbo em 1ª p.p. na oração anterior

(114) *Fizemos* um bingo aqui assim na ci... no barrio em prol mesmo do próprio times mesmo que **a gente** tem

l) com verbo em 3ª p. p. na oração anterior

(115) Aí, eu acho bom... nem nada, num chegô até hoje. Tem mais de um mês que *fo-*
ram medi e num **chegaro** nada ainda.

m) com categoria vazia na oração anterior

(116) Vai, faz o quê? \emptyset *Raspa* a mandioca, \emptyset **faz** como já expliquei aqui a senhora e tal...

3.2.6 Variáveis Sociais

3.2.6.1 Localidade do falante

A variação diatópica é um fator importante para que possam ser interpretados resultados obtidos em pesquisas lingüísticas. Alguns fenômenos lingüísticos podem ser percebidos com maior intensidade em determinadas regiões do país, e quanto mais isolada for uma comunidade de fala, mais conservador será o seu caráter. Os falantes das comunidades afro-brasileiras, por exemplo, devido ao seu isolamento geográfico e ao difícil acesso aos meios de comunicação e à escolarização, mantêm marcas lingüísticas conservadoras. Segundo Weirich, Labov e Herzog (2006 [1968]),

O problema de explicar a transição geográfica de dialetos através de um território parece, portanto, simétrico ao problema de explicar a transição de dialetos através do tempo numa comunidade. Em cada caso, há um contato entre falantes com sistemas diferentes.

Na codificação das ocorrências, foram mantidas as seis localidades das quais os informantes precedem: as comunidades de Helvécia, Rio de Contas, Sapé e Cinzento; a zona rural ou a sede do município de Santo Antonio de Jesus. Isso se deu para que, se houvesse comportamentos lingüísticos próprios de algumas localidades que pudessem enviesar os dados, isto fosse facilmente identificável.

Posteriormente, os dados provenientes das comunidades afro-brasileiras foram agrupados sob uma só variante. Desta forma, na variável social *localidade do falante*, são analisadas três variantes: comunidades afro-brasileiras, zona rural e a sede do município de Santo Antonio de Jesus.

3.2.6.2 Gênero

Desde as primeiras referências aos comportamentos lingüísticos de homens e mulheres, a variável gênero tem sido foco de diversos estudos. Fischer (1958 Apud PAIVA, 2003) é quem primeiro estabelece uma correlação entre variação lingüística e o gênero do falante e constata que as mulheres tendem a utilizar mais em sua fala as formas de prestígio. Esta é, no que concerne à variável gênero, uma das descobertas de maior relevância das investigações sociolingüísticas (FERNÁNDEZ, 1998).

Quanto ao papel da variável gênero, entretanto, os resultados dos estudos já realizados mostram não haver ainda conclusões definitivas sobre o assunto, visto as diversas tendências para a fala de homens e mulheres observadas até hoje, além do mais, a distinção entre variante de prestígio e variante estigmatizada nem sempre é tão nítida, ou, simplesmente, inexistente. Conforme Paiva (2003, p. 36)

No estudo da correlação gênero/sexo e mudança lingüística, um aspecto a considerar é o valor social da variante inovadora. Um processo de mudança pode ser a instalação de uma forma prestigiada socialmente ou de uma forma estigmatizada, que infringe padrões lingüísticos vigentes. A distinção entre esses dois tipos de mudança permite definir com maior clareza o papel da variável gênero/sexo nos processos de mudança.

Desta forma, não se poderia pensar em resultados conclusivos acerca da mudança lingüística baseados somente no gênero do falante. Para Fernández (1998, p. 35), “El sexo puede mostrarse, por tanto, más como un factor de segundo orden, como algo que suele subordinarse a dimensiones sociales diferentes y con mayor poder de determinación.” O comportamento lingüístico de homens e mulheres poderia ser influenciado, por exemplo, pela localidade do informante.

Os resultados da variável social gênero em Antonino (2007) mostram que implementação da regra de concordância de gênero nos predicativos do sujeito e em estruturas passivas no português popular do estado da Bahia é liderada por mulheres – que alcançam uma frequência de uso da concordância de 8%, enquanto os homens obtêm apenas 1%. Embora não classificadas como grandes centros urbanos, as cidades de Poções e Santo Antonio de Jesus, acerca das quais se desenvolve o citado estudo, vem passando por um processo recente de expansão. Antonino tece as seguintes considerações acerca do comportamento social das mulheres no interior do estado:

É nítido que a realidade social do interior do país vem mudando; a mulher não está mais confinada a cozinha de sua casa, não é mais apenas mãe. Ela é trabalhadora, muitas vezes responsável pelo sustento do lar, saindo de casa tanto quanto ou mais que os homens, saindo inclusive de sua cidade em busca de melhoria de sua qualidade de vida.

Em comunidades isoladas, pelo fato de os homens terem mais contato com o exterior, faz com que geralmente liderem o uso de variantes padrão. R. Cavalcante (2007) aborda as estratégias de negação sentencial no vernáculo de comunidades afro-brasileiras isoladas – as entrevistas analisadas compõem o *Corpus Base do Português Afro-brasileiro do Estado da*

Bahia do Projeto Vertentes. Os resultados percentuais demonstram que os homens alcançam maior frequência do que as mulheres quando se trata da variante padrão, a negação pré-verbal – em 69% dos enunciados produzidos por homens a negação era pré-verbal, enquanto as mulheres atingem o índice de 65%. Quando são analisadas as frequências referentes à estratégia não-padrão [Não V não], os resultados se invertem: os enunciados produzidos por as mulheres totalizam 31% das ocorrências contra 24% dos homens.

3.2.6.3 Faixa etária

O conhecimento do comportamento lingüístico dos falantes em função da sua faixa etária permite afirmar se um dado fenômeno variável está em processo de implementação de uma variante ou se pode ser classificado como um caso de variação estável. O estudo da faixa etária permite, ainda, comprovar a existência de variantes conservadoras ou inovadoras.

La edad de los hablantes [...] es uno de los factores sociales que con mayor fuerza e claridad pueden determinar los usos lingüísticos de una comunidade de habla. En cierto modo, puede afirmarse que que la edad condiciona la variación lingüística con más intensidad que otros factores, también importantes como el sexo o la classe social. (FERNÁNDEZ, 1998, p. 40)

Normalmente, falantes de faixas etárias mais avançadas tendem a preferir formas mais conservadoras, enquanto falantes mais jovens preferem formas inovadoras, e isto pode ser observado, sobretudo, no nível lexical. Entretanto, esta diferença não interfere na comunicação entre falantes de faixas etárias distintas, pois todos os falantes são capazes de entender e utilizar todas as formas (NARO, 2003, p. 44).

As faixas etárias pelas quais se distribuem os falantes neste estudo são três:

- Faixa I – de 20 a 40 anos
- Faixa II – de 40 a 60 anos
- Faixa III – mais de 60 anos

3.2.6.4 Estada fora da comunidade

Assim como é relevante saber origem do falante, também é importante o controle da influência que este pode receber através do contato com outras comunidades de fala. Baxter e

Lucchesi (1995), em estudo sobre a variação na concordância verbal com a 1ª pessoa do singular, verificaram a influência do exterior na fala de informantes através da variável social estada fora da comunidade por seis meses ou mais. Os informantes que haviam morado fora da comunidade lideram o processo de implementação da regra de concordância e atingem índices de 93% de concordância, enquanto os que nunca saíram da comunidade obtêm 75%. Os pesos relativos alcançados são de, respectivamente, .67 e .41. A variável permite, portanto verificar a influência do exterior na fala dos informantes.

No presente estudo, a variável estada fora da comunidade distribui os indivíduos por dois grupos: os que já viveram fora da comunidade por pelo menos seis meses e aqueles que nunca se ausentaram da comunidade por tempo superior a este período.

3.2.6.5 Escolaridade

O grau de escolarização de um indivíduo tem se mostrado uma variável muito produtiva nos estudos sociolingüísticos. Esta variável é comumente utilizada para que possa ser verificada a sua influência quanto à apropriação de formas de prestígio pelos falantes. (MOLLICA, 2003, p. 28).

Conforme Fernández (1998, p.55)

La sociolingüística [...] há comprobado que el nivel educativo de los hablantes determina de forma directa y clara la variación linguística: es normal que las personas más instruidas hagan mayor uso de las variantes que son consideradas como más prestigiosas o que más se ajustan a la norma.

Como citado anteriormente, todos os informantes que compõem esta pesquisa possuem pouca ou nenhuma escolaridade e, aqui, se distribuem entre analfabetos ou semi-analfabetos. Segundo a metodologia do *Projeto Vertentes*, tem-se como semi-analfabetos aqueles falantes que tiveram uma experiência ao menos mínima com a escola. Incluímos neste grupo também aqueles que apenas assinam o nome.

A hipótese acerca da escolaridade, neste trabalho, é a de que na fala de informantes que tenham tido algum contato com a escola sejam mais frequentes as formas de indeterminação como $\emptyset + 3VPP$, atestada pelas gramáticas tradicionais.

4 A ANÁLISE DOS RESULTADOS

Ao fim da análise quantitativa realizada para o presente trabalho, foi obtido um total de 3.830 ocorrências de indeterminação do sujeito. As duas estratégias que obtiveram a maior frequência foram o pronome *a gente*, com 47% das ocorrências, seguido da forma $\emptyset + V3PS$, com 25%. A terceira estratégia mais utilizada pelos falantes do português popular do interior do estado da Bahia foi o pronome *você*, que obteve nove pontos percentuais. Das estratégias restantes nenhuma ultrapassou o índice de 6%. Portanto, a distância entre a frequência das duas primeiras estratégias citadas e as demais é notável. Os resultados desta primeira análise se distribuem da seguinte forma:

Tabela 7: Distribuição geral das estratégias de indeterminação do sujeito no português popular do interior do estado da Bahia.

Estratégia de Indeterminação	N	%
A gente	1789	47
$\emptyset + V3PS$	963	25
Você	347	9
Nós	225	6
Eles	221	5,7
$\emptyset + V3PP$	161	4,2
$\emptyset + V1PP$	47	1,2
Ele	41	1
O senhor	23	0,6
Eu	13	0,3
TOTAL	3.830	100

Observou-se primeiramente o fenômeno da realização fonética do sujeito pronominal, na medida em que vários dos sujeitos pronominais passaram a ser empregados como estratégia para expressar a indeterminação do sujeito. Nesse sentido, é importante observar que as formas pronominais são responsáveis por quase 70% das ocorrências de indeterminação no interior do estado da Bahia. Duarte (1995), ao verificar a perda do princípio “Evite Pronome” (Cf. Chomsky, 1981) na norma urbana culta carioca da década de 90, chega à conclusão de que as mudanças que afetam o sujeito de referência definida afetariam também o sujeito de referência arbitrária, ou seja, o sujeito indeterminado. Isto pode ser constatado, sobretudo quando são analisados os índices de realização destes sujeitos. No português popular do inte-

rior do estado da Bahia não é diferente: das 2.622 ocorrências de estratégias pronominais obtidas nesta pesquisa – com exceção das forma *eu* e *senhor*, excluídas da análise – 2.117 foram realizadas foneticamente, 80% do total, conforme pode ser visto na tabela 8.

Tabela 8: A indeterminação do sujeito segundo a variável *realização fonética do sujeito indeterminado*.

Estratégia	Realização	
	N	%
A gente	1411 / 1789	79
Você	291 / 346	84
Nós	194 / 225	86
Eles	189 / 221	85
Ele	32 / 41	78
TOTAL	2117 / 2622	80

Com base nesse primeiro escrutínio do fenômeno variável, algumas estratégias quantificadas foram agrupadas, após uma análise dos seus resultados quantitativos. Desta forma, foram excluídos os pronomes *eu* e *ele*. O número de ocorrências obtidas pelo pronome *eu* - que totalizou treze apenas - tornaria incertas as conclusões que pudessem ser feitas sobre o seu comportamento. Quanto ao pronome *ele*, este foi excluído desta análise pelo baixo número de ocorrências e também por não ter sido possível constatar a produção de referências externas à fala, tais como sinais, concomitantemente à utilização da estratégia, o que invalidaria a classificação da ocorrência como pertencente à esfera da indeterminação. As ocorrências do pronome de tratamento *senhor(a)* foram unidas às ocorrências do pronome *você*, verificadas as propriedades semelhantes destes. O pronome *nós* e as ocorrências da construção \emptyset + *VIPP* foram agrupadas sob a forma *nós*.

As demais estratégias não sofreram alterações quanto ao seu número de ocorrências nem foram agrupadas – o pronome *eles* e a forma \emptyset + *V3PP* foram mantidos como estratégias distintas a fim de se verificar o comportamento da única estratégia padrão presente na gramática das comunidades de fala analisadas, visto que a outra estratégia prevista pelas GTs, o verbo em 3ª p. s. acompanhado pela partícula *se*, não foi analisada quantitativamente por não ter alcançado um número de ocorrências que o permitisse²⁸. Assim, as estratégias que tiveram o seu comportamento descrito nesta análise são as seguintes:

²⁸ As poucas ocorrências da partícula *se* com valor de indeterminação do sujeito serão analisadas qualitativamente ao final deste capítulo, na seção 4.3.

Tabela 9: Estratégias de indeterminação do sujeito.

Estratégia de Indeterminação	N	%
A gente	1789	47
Ø + V3PS	963	26
Você	369	10
Nós	272	7
Eles	221	6
Ø + V3PP	161	4
TOTAL	3775	100%

A seguir, serão apresentados os resultados da distribuição das estratégias de indeterminação pelos níveis de referencialidade do sujeito indeterminado.

4.1 OS NÍVEIS DE REFERENCIALIDADE DO SUJEITO INDETERMINADO

Conforme foi dito no segundo capítulo, as estratégias de indeterminação são distribuídas, nesta análise, por quatro níveis de referencialidade do sujeito indeterminado – o nível *genérico*, de caráter universal, o nível *grupo parcialmente definido*, o *eu ampliado* e o nível do *agente específico indefinido*. O maior número de ocorrências foi encontrado no nível *genérico*, com 41% do total. O nível *grupo parcialmente definido* aparece logo em seguida, atingindo 32% de frequência. Os níveis *eu ampliado* e *agente específico indefinido* obtiveram, respectivamente, 14 e 13 por cento de uso. A distribuição das estratégias de indeterminação do sujeito pelos seus níveis de referencialidade é apresentada na tabela 10.

Tabela 10: Distribuição das estratégias de indeterminação do sujeito pelos seus níveis de referencialidade.

Nível de Referencialidade	Estratégia												T
	A gente		Nós		Você		Eles		Ø + V3PP		Ø + V3PS		
	N	%	N	%	N	%	N	%	N	%	N	%	
Genérico	561	36	15	1	365	24	21	1	3	-	582	38	1547
Parcialmente definido	796	66	188	16	-	-	15	1	3	-	202	17	1204
Eu ampliado	432	79	69	13	-	-	-	-	1	-	41	8	543
Específico indefinido	-	-	-	-	4	1	185	38	154	32	138	29	481
TOTAL	1789	47	272	7	369	10	221	6	161	4	963	26	3775

As estratégias de indeterminação do sujeito se distribuem pelos níveis de referencialidade analisados de modo que podem ser constatadas três estratégias principais para cada um desses níveis.

No nível *genérico*, os pronomes *a gente* e *você* e a forma $\emptyset + V3PS$, que alcançam, respectivamente, 36, 24 e 38 pontos percentuais, totalizam 98% das ocorrências deste nível. Nos níveis *grupo parcialmente definido* e *eu ampliado*, predominam as mesmas estratégias: os pronomes de primeira pessoa do plural *a gente* e *nós* e a forma $\emptyset + V3PS$. As demais estratégias, nestes níveis, não ultrapassam 1%. Quanto ao nível *agente específico indefinido*, observam-se as estratégias de terceira pessoa do plural – o pronome *eles* e a construção $\emptyset + V3PP$ – e, mais uma vez, $\emptyset + V3PS$.

Nas sessões apresentadas a seguir, serão detalhados os comportamentos das estratégias que se revezam pelos citados níveis de referencialidade, bem como será explicada a ocorrência, em cada um dos níveis, de demais estratégias.

O fato de $\emptyset + V3PS$, uma estratégia menos marcada, ter sido a única a se distribuir pelos quatro níveis de referencialidade aqui estudados contribui com as hipóteses acerca dos processos que balizaram a aquisição de língua portuguesa e que ainda se refletem em nossa estrutura lingüística.

4.1.1 Nível genérico

As ocorrências dos pronomes *nós* e *eles* e da forma $\emptyset + V3PP$ foram excluídas desta parte da análise pelas suas baixas frequências, que totalizaram 2%. Contudo chamou-nos atenção o fato de o pronome *eles* ter 21 ocorrências no nível genérico, número superior ao alcançado pelo pronome *nós*, com 15 ocorrências (cf exemplo 115). As ocorrências do pronome *eles* foram, então, analisadas separadamente, localizadas com o programa TSORT, que compõe o pacote de programas VARBRUL, e, assim, constatou-se que são provenientes da fala de apenas cinco dos 48 informantes cujas entrevistas compõem o *corpus* dessa pesquisa. As ocorrências referem-se a procedimentos próprios da comunidade (116) ou a situações hipotéticas (117) em que o sujeito *eles* não possui nenhuma referência anterior e em que poderia, seguramente, ser substituído pela partícula *se*, de caráter universalizante.

(115) É aonde *nós* tamos por aí até o dia que Deus querê.

(116) DOC: Hum... E que remédio é esse que dá assim?

INF: *Eles* dão o tá da 'um moscada... \emptyset rala, \emptyset coloca na água morna, \emptyset dá pra bebê...

(117) *Eles* constrói a violência e \emptyset constrói a felicidade do ôtro lado.

As três estratégias que obtiveram os maiores percentuais neste nível de referencialidade foram os pronomes *a gente* e *você* e a forma $\emptyset + V3PS$, que se distribuíram de forma homogênea pelo nível genérico. As estratégias *a gente* e $\emptyset + V3PS$ obtiveram percentuais praticamente idênticos de uso, 38% e 39% respectivamente. A estratégia *você* alcançou 23 pontos.

Tabela 11: Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade *genérico*.

Estratégia	N	%
A gente	561	38
Você	342	23
$\emptyset + V3PS$	582	39
TOTAL	1485	100

Somente nesse nível de referencialidade ocorre uma distribuição equilibrada de três estratégias claramente distintas de indeterminação do sujeito: o pronome da primeira pessoa do plural, *a gente*; o pronome da segunda pessoa do singular, *você*; e a forma verbal não marcada da terceira pessoa do singular. Nos demais níveis de referencialidade, como poderá ser visto a seguir, as variantes de uma mesma estratégia de indeterminação acabam prevalecendo.

4.1.2 Grupo parcialmente definido

As estratégias que prevaleceram no nível *grupo parcialmente definido* foram os pronomes de primeira pessoa do plural, *a gente* e *nós*, e a forma $\emptyset + V3PS$. Foram excluídas da análise desse nível de referencialidade as ocorrências de *eles* e de $\emptyset + V3PS$, que juntas não alcançaram 2% de frequência. O exemplo (118) apresenta ocorrências do pronome *eles* classificadas como pertencentes ao nível de referencialidade do *grupo parcialmente definido*. Nesse contexto o pronome *eles* não possui referência direta, porém é fácil perceber que a ocorrência se refere às pessoas da família do informante pelo uso do termo “den’de casa”.

(118) Pra isso eu... até hoje eu sô assim, eu chego den’de casa, eu num peço... se precisá de uma coisa pra comprá, se eu tivé dinhêro eu compro, seu eu num tivé, eu

num peço a ninguém pra comprá. *Eles* que espie se precisa ou não, mas pra eu dizê assim “falta comprá um açúcar, um café”... Se tivé com dinhêro no bolso eu vô e compro, boto den’de casa, agora se num tivé, eu num... Pra eu chegá assim “óia, vai comprá um açúcar, um café”, não. *Eles* ‘tá vendo que falta café e açúcar, *eles* compre.

O pronome *a gente* prevalece no nível de referencialidade *grupo parcialmente definido*: obteve 67% de freqüência. A forma $\emptyset + V3PS$ e o pronome *nós* obtiveram 17% e 16%.

Tabela 12: Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade *grupo parcialmente definido*.

Estratégia	N	%
A gente	796	67
Nós	188	16
$\emptyset + V3PS$	202	17
TOTAL	1186	100

A soma das porcentagens atingidas pelas variantes da primeira pessoa do plural totaliza mais de oitenta por cento do total de ocorrências nesse nível de referencialidade (precisamente 83%). Constata-se, dessa forma, que os pronomes da 1ª pessoa do plural são estratégias que dominam amplamente indeterminação do sujeito nos casos em que a referência é um grupo de contorno pouco definidos.

4.1.3 Eu ampliado

Apenas uma ocorrência da forma $\emptyset + V3PP$ deste nível foi excluída para que pudéssemos analisar o comportamento das três estratégias que prevaleceram. Esta ocorrência é apresentada no exemplo (119).

(119) DOC: Hum. E o gado aqui come na capoeira mesmo, é?

INF: Hum, hum... é.

DOC: Ôcê, às vezes... dá alguma coisa assim tipo ração, alguma coisa...?

INF: É... às vez **dão**, né...

Nesta ocorrência, mesmo após o documentador fazer a pergunta usando a 2ª pessoa indireta, o *você*, o informante utilizou-se da flexão verbal de 3ª pessoa do plural para se referir a um costume próprio da sua comunidade, em que ele mesmo se incluía.

Os resultados da distribuição das formas *a gente*, *nós* e $\emptyset + V3PS$ podem ser vistos na tabela 13.

Tabela 13: Distribuição das estratégias de indeterminação do sujeito pelo nível de referencialidade *eu ampliado*

Estratégia	N	%
A gente	432	79
Nós	69	13
$\emptyset + V3PS$	41	8
TOTAL	542	100

Nesse nível de referencialidade, a preferência por uma das formas do pronome da 1ª pessoa do plural ainda é maior, pois, reunidas, as ocorrências de *nós* e *a gente* correspondem a mais de noventa por cento do total (precisamente 92%).

4.1.4 Agente específico indefinido

As quatro ocorrências do pronome *você* classificadas como pertencentes ao nível de referencialidade do *agente específico indefinido*, retiradas da análise deste nível, foram localizadas em um só turno da entrevista de um informante da comunidade de Sapé, que integra as comunidades afro-brasileiras aqui estudadas. A ocorrência é apresentada no exemplo (120).

(120) E peda...e peda...e peda...e peda, pá gente i(r) em Coração tava assim...e só ia assim, pra lá gruta, pra cá gruta, pra li gruta, num sei onde gruta... gente andava nessa peda que saia na de lá e ia embora...bem... e subiu a ladêra, discia ôta... subia uma ladêra e discia ôta, o que...levava um...um...vela, candiêro, faço de...de...de bagaço de cana...té, graças a Deus, saia fora... já agora que eu fui, eu num vi coração, dismantelô ondê coração tava...a gruta que tinha, **cês** dismantelô... ôta que tinha, **cês** mantelô, assim ondê que tinha, **cês** mantelô e **fez** uma igreja [...].

Nessa ocorrência, o falante utiliza-se do pronome *vocês* para referir-se a um grupo de pessoas, em sua opinião, responsáveis pela destruição de uma parte da gruta à qual seguiam em romaria.

As demais variáveis dependentes se distribuíram pelo nível do *agente específico indefinido* da seguinte forma:

Tabela 14: Distribuição das estratégias de indeterminação do sujeito pelo nível *agente específico indefinido*.

Estratégia	N	%
Eles	185	39
Ø + V3PP	154	32
Ø + V3PS	138	29
TOTAL	477	100

Nesse caso, o pronome da 3ª pessoa do plural, realizado foneticamente ou não, é estratégia preferencial de referência a um agente específico indefinido, correspondendo a mais de setenta por cento do total de ocorrências nesse nível de referencialidade (precisamente 71%).

4.2 A INDETERMINAÇÃO DO SUJEITO NO INTERIOR DO ESTADO DA BAHIA

Nesta seção serão descritos os comportamentos das seis estratégias referidas no início deste capítulo: os pronomes *a gente*, *nós*, *you* e *eles* e as formas Ø + V3PP e Ø + V3PS. Para a obtenção dos pesos relativos, cada estratégia foi confrontada com as demais. Dessa forma, são apresentadas as variáveis que, segundo o programa VARBRUL, são relevantes estatisticamente e favorecem a escolha de um ou de outro valor da variável dependente.

4.2.1 A gente

A forma *a gente* foi a primeira estratégia em número de ocorrências nessa análise: 47% do total. Com relação a esta variante, o programa VARBRUL não considerou como relevantes sete variáveis: realização do sujeito, desinência verbal, tipo de verbo, tipo de frase, modo verbal, gênero e escolaridade. A variável *inclusão do falante* foi excluída desta análise por não haver ocorrências do pronome *a gente* em função indeterminadora que não tivessem o falante incluso, o que deixa a variável com um único fator e impossibilita a sua análise.

As variáveis selecionadas foram cinco, apresentadas a seguir segundo o seu grau de importância.

Tabela 15: Variáveis selecionadas na análise da estratégia *a gente*

Variável	
1°	Forma antecedente
2°	Nível de referencialidade
3°	Localidade do falante
4°	Faixa etária
5°	Estada fora

Os resultados do pronome *a gente* quanto à forma antecedente são apresentados na tabela 16. Sob a nomenclatura *outras formas* estão as ocorrências de cinco estratégias antecedidas pelo pronome *a gente*: *você*, $\emptyset + V3PS$, *ele*, *eles* e $\emptyset + V3PP$.

Tabela 16: O pronome *a gente* segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	1188 / 2153	55	.47
A gente	545 / 581	94	.92
$\emptyset + V3PS$	32 / 335	10	.08
Nós	8 / 40	20	.10
Outras formas	16 / 132	12	.14
TOTAL	1789 / 3241	55	

Nível de significância .000

O pronome *a gente*, em 94% das ocorrências, foi antecedido pela mesma forma, realizada ou não. O peso relativo desta variante foi de **.92** – o que demonstra que, uma vez utilizada a estratégia, a sua manutenção é praticamente certa. Esses resultados são semelhantes aos verificados em Cunha (1993) e Santana (2006), em que a variável *forma antecedente* também foi selecionada como o primeiro grupo favorecedor do pronome *a gente* indeterminador; os pesos relativos obtidos nos dois estudos citados foram de .93. O fator 1ª menção se mostrou neutro quanto à sua porcentagem, mas o peso relativo mostrou ser esse um fator levemente desfavorecedor da estratégia *a gente*.

Não foram encontradas ocorrências do pronome *a gente* no nível de referencialidade

do *agente específico indefinido*; estas se distribuíram pelos três níveis restantes sendo predominantes nos níveis *eu ampliado* e *grupo parcialmente definido*.

Tabela 17: O pronome *a gente* segundo a variável *grau de indeterminação*.

Nível de referencialidade	N / T	%	P. R.
Genérico	561 / 1501	37	.31
Parcialmente definido	796 / 1197	66	.63
Eu ampliado	432 / 544	80	.74
TOTAL	1789 / 3241	55	

Nível de significância .000

O pronome *a gente*, além de ter sido mais freqüente no nível *eu ampliado*, em que foi responsável por 79% das ocorrências, obteve nesse nível o peso relativo de **.74**. Logo em seguida, como favorecedor da estratégia, aparece o nível do *grupo parcialmente definido*, com peso relativo de **.63**. O nível *genérico*, ao contrário, desfavorece o uso do pronome. A predominância da estratégia *a gente* nos níveis *eu ampliado* e *grupo parcialmente definido* se deve, provavelmente, ao traço semântico desse pronome, que abrange inevitavelmente o falante. Como já mencionado na análise da tabela 10, não houve ocorrência do pronome *a gente* no nível do *agente específico indefinido*.

A *localidade do informante* foi a terceira variável selecionada como relevante para o uso do pronome *a gente*; as freqüências obtidas pelas suas variantes constam na tabela 18.

Tabela 18: O pronome *a gente* segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Com. Afro-brasileira	791 / 1720	46	.42
Zona Rural	464 / 748	62	.58
Zona Urbana	534 / 773	69	.60
TOTAL	1789 / 3241	55	

Nível de significância .000

Nas comunidades afro-brasileiras isoladas, o pronome *a gente* é responsável por 46% das ocorrências de indeterminação do sujeito, a freqüência aumenta 14 pontos na zona rural, e atinge quase 70% na sede do município. Esse contínuo se manteve quando analisados os pesos relativos: o pronome *a gente* é desfavorecido nas comunidades afro-brasileiras, com **.42** de peso relativo, e é favorecido tanto na zona rural quanto na sede da cidade, em que a-

tinge, respectivamente, os índices de ,58 e ,60. Os resultados apontam para a implementação da estratégia indeterminadora *a gente* a partir do pólo mais urbanizado deste estudo: a sede da cidade de Santo Antonio de Jesus.

Os resultados da variável social *faixa etária* são apresentados a seguir.

Tabela 19: O pronome *a gente* segundo a variável social *faixa etária*.

Faixa etária do informante	N / T	%	P. R.
Faixa I	753 / 1238	61	.55
Faixa II	719 / 1225	59	.51
Faixa III	317 / 778	41	.40
TOTAL	1789 / 3241	55	

Nível de significância .000

Os níveis alcançados pelas faixas I, II e III demonstram a implementação do pronome *a gente*, que alcança seu maior peso relativo, ,55, na fala de informantes da faixa I – de 20 a 40 anos. À faixa II é atribuído o peso relativo de ,51, o que demonstra que o fator é praticamente neutro. O peso relativo e a frequência de uso do pronome na faixa III demonstram que falantes dessa faixa etária, de 60 anos em diante, utilizam pouco o pronome *a gente*, provavelmente em detrimento das variantes mais conservadoras. Verificados os resultados das variáveis sociais *localidade* e *faixa etária*, foi feito um cruzamento dessas a fim de que pudéssemos analisar o percurso da implementação do pronome *a gente* no português do interior do estado da Bahia. Os dados da rodada submetida ao VARBRUL são apresentados na tabela 20.

Tabela 20: O pronome *a gente* segundo o cruzamento das variáveis sociais *localidade* e *faixa etária*.

Localidade	Faixa etária	N / T	%	P. R.
Sede	I	191 / 251	76	.69
	II	255 / 387	66	.59
	III	88 / 135	65	.52
Zona Rural	I	104 / 205	51	.48
	II	286 / 406	70	.65
	III	74 / 137	54	.57
Afro-brasileira	I	394 / 646	61	.55
	II	242 / 568	43	.38
	III	155 / 506	31	.29
TOTAL		1789 / 3241	55	

Nível de significância: 006

Os resultados do cruzamento dessas duas importantes variáveis sociais configuram um paralelo com o processo da implementação do *a gente*, em detrimento do *nós*, nos sujeitos com referência definida no português brasileiro e apontam fortemente para um cenário de mudança em curso no sentido da implementação do *a gente* como estratégia de indeterminação do sujeito. Esse processo de implementação é mais nítido na zona urbana e nas comunidades rurais afro-brasileiras isoladas, observando-se uma certa flutuação na zona rural do município de Santo Antônio de Jesus. Os maiores percentuais de uso do pronome *a gente* foram encontrados na sede do município: 76%, 66% e 65% para as faixas I, II e III respectivamente. Esses dados revelam que a zona urbana lidera a mudança em favor da forma inovadora e também confirmam a direção da mudança – o uso crescente dessa estratégia na medida em que se passa para as gerações mais novas. Na zona rural da cidade, a estratégia demonstra-se mais produtiva na fala dos informantes da faixa etária intermediária, configurando um quadro de variação estável, que foge ao padrão mais geral. E, nas comunidades afro-brasileiras isoladas, repete-se o padrão observado na comunidade de fala da zona urbana: os maiores índices são alcançados na fala da geração mais nova e em que os valores decrescem à medida que se passa para a fala dos mais velhos.

Portanto, não obstante o desvio observado na zona rural, há resultados empíricos que podem satisfatoriamente fundamentar uma generalização no sentido da implementação do *a gente* como estratégia de indeterminação no português popular do interior do Estado da Bahia, sendo que os núcleos urbanos do interior do Estado liderariam o processo de mudança, que, em seguida, se difundiria pela zona rural, atingindo inclusive as comunidades rurais mais isoladas.

O último fator selecionado foi a variável social *estada fora da comunidade*, representada na tabela 21.

Tabela 21: O pronome *a gente* segundo a variável social *estada fora da comunidade*.

Estada fora da comunidade	N / T	%	P. R.
Sim	928 / 1516	61	.54
Não	861 / 1725	50	.46
TOTAL	1789 / 3241	55	

Nível de significância .000

Os dados relativos tanto à frequência quanto ao peso relativo do pronome *a gente* nesta variável indicam que a saída da comunidade por seis meses ou mais favorece o uso do pro-

nome *a gente*. Enquanto os falantes que saíram da comunidade atingem um peso relativo de ,54, os que não saíram atingem ,46. Esses resultados ratificam a interpretação aqui adotada de implementação do *a gente* como estratégia de indeterminação do sujeito no interior do país a partir da influência dos grandes centros urbanos (cf. Lucchesi, 2001 e 2006). Dessa forma, os indivíduos que se ausentaram de suas comunidades de fala de origem, por pelo menos seis meses, geralmente para trabalhar nos grandes centros urbanos, são exatamente aqueles que lideram esse processo de mudança.

4.2.2 Nós

De acordo com os resultados fornecidos pelo programa *VARBRUL*, o pronome *nós* é favorecido por sete das treze variáveis que compunham a chave de codificação utilizada nesta pesquisa, dispostas segundo sua relevância na tabela 22.

Tabela 22: Variáveis favorecedoras da estratégia *nós*

Variável	
1°	Nível de referencialidade
2°	Forma antecedente
3°	Localidade do falante
4°	Faixa etária
5°	Modo verbal
6°	Gênero
7°	Escolaridade

Assim como o que foi verificado na análise do pronome *a gente*, o pronome *nós* não obteve ocorrências no nível de referencialidade do *agente específico indefinido*, como é mostrado na tabela 23.

Tabela 23: O pronome *nós* segundo a variável *nível de referencialidade*.

Nível de referencialidade	N / T	%	P. R.
Genérico	15 / 1547	1	.21
Parcialmente definido	188 / 1204	16	.76
Eu ampliado	69 / 543	13	.77
TOTAL	272 / 3294	8	

Nível de significância .010

Os níveis em que a representatividade do pronome *nós* foi maior – o *eu ampliado* e o *grupo parcialmente definido* – obtiveram pesos relativos praticamente idênticos: ,77 e ,76. Sua distribuição pelo nível *genérico* foi mínima, e o peso relativo alcançado foi ,21.

A variável *forma antecedente* foi a segunda selecionada pelo programa *VARBRUL* como estatisticamente relevante. Quanto aos fatores que compõem essa variável, só foram encontradas ocorrências de três formas antecedidas pelo pronome *nós*: *a gente*, \emptyset + *V3PS* e *senhor*, unidas sob a nomenclatura *outras formas*.

Tabela 24: O pronome *nós* segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	206 / 2153	10	.60
Nós	44 / 56	79	.97
Outras formas	22 / 1085	2	.28
TOTAL	272 / 3294	8	

Nível de significância .010

Quanto às formas antecedidas pelo pronome *nós*, 79% eram do próprio pronome. O peso relativo do pronome *nós* como forma antecedente é de ,97, o que demonstra uma aplicação praticamente categórica. O contexto de *primeira menção*, com peso relativo de ,60, indica um ligeiro favorecimento a essa variante. Quando antecipada pelas demais formas de indeterminação do sujeito, a frequência dessa variante cai para um peso relativo de ,28. Portanto, os resultados dessa variável confirmam o *princípio do paralelismo formal* (cf. 3.2.5.8).

A *localidade do informante*, assim como o observado na análise da forma *a gente*, foi a terceira variável selecionada como relevante para o uso do pronome *nós*, porém com resultados simetricamente opostos.

Tabela 25: O pronome *nós* segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Com. Afro-brasileira	199 / 1751	11	.63
Zona Rural	44 / 755	6	.43
Zona Urbana	29 / 788	4	.28
TOTAL	272 / 3294	8	

Nível de significância .010

Esse cenário é precisamente complementar ao observado com o processo de implementação do *a gente*. Isso demonstra que o *a gente* está sendo implementado como estratégia de indeterminação em detrimento do *nós*. Onde o uso do *a gente* está mais avançado (na zona urbana), o uso do *nós* é mais restrito; e, onde a implementação do *a gente* está mais atrasada (nas comunidades rurais afro-brasileira), o uso do *nós* é mais freqüente. A zona rural do município estaria numa situação intermediária.

Quanto à variável social *faixa etária*, selecionada pelo programa *VARBRUL* em quarto lugar, os resultados são os seguintes:

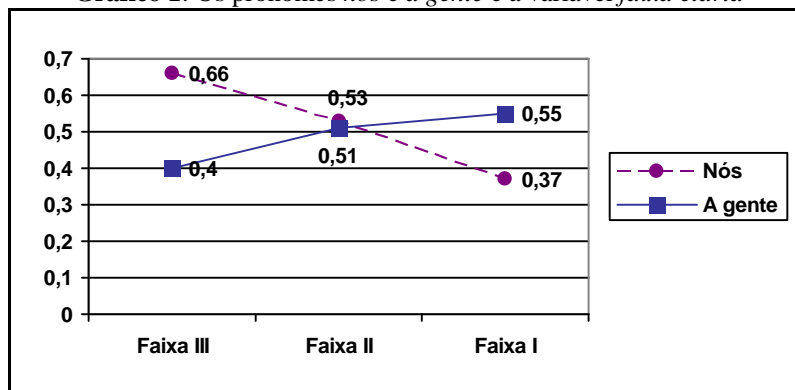
Tabela 26: O pronome *nós* segundo a variável *faixa etária*

Faixa etária	N / T	%	P. R.
Faixa I	72 / 1256	6	.37
Faixa II	104 / 1252	8	.53
Faixa III	96 / 786	12	.66
TOTAL	272 / 3294	8	

Nível de significância .010

Aqui também se constata um quadro complementar ao observado para o *a gente*. Se a distribuição geracional indica uma mudança em progresso em favor do *a gente* (cf. tabela 19), essa mesma distribuição geracional indica uma mudança no sentido do desaparecimento da forma *nós*, já que seu uso diminui progressivamente na medida em que se passa da fala dos mais velhos para os mais jovens (como se pode ver no gráfico abaixo). Todo esse cenário deixa claro que, no processo de mudança, o *a gente* vem substituindo o *nós*.

Gráfico 1: Os pronomes *nós* e *a gente* e a variável *faixa etária*



A variável *modo verbal* também se demonstrou relevante quanto ao uso do pronome *nós*. Os seus resultados constam na tabela 27.

Tabela 27: O pronome *nós* segundo a variável *modo verbal*.

Aspecto	N / T	%	P. R.
Realis	248 / 2064	12	.58
Irrealis	24 / 1230	2	.36
TOTAL	272 / 3294	8	

Nível de significância .010

As ocorrências do pronome *nós* são favorecidas, tanto no percentual quanto no peso relativo (**,58**), pelo modo verbal *realis*. Essas ocorrências são exemplificadas no item (121), em que o informante utiliza os pronomes *nós* e *a gente* durante o turno em que se refere a fatos passados de sua vida, circunscritos, portanto, ao modo *realis*. Já o uso da estratégia em questão nas orações de modo verbal *irrealis* é desfavorecido: o peso relativo da variante quanto ao uso do pronome *nós* é de **,36**. No exemplo (122), o informante, durante sua narração, supõe que poderia ter agido diferente quanto ao cuidado que teve com a máquina de mandioca – faz, portanto, ainda que não tenha concluído o seu pensamento, uma suposição.

(121) A gente num tinha trabalho... Meu pai bebia muito na época. Minha mãe também. Então a gente vivia mesmo praticamente... é... abandonada. Depois é que a gente foi começano a reeguê nossa vida. *Nós* tamo nessa casa hoje derna que a gente... Eu... que eu vim pra aqui...

(122) Nós fazia, que nós tinha uma mandioca... um ININT manual... tinha não, temos! Só que a nossa tá parada, porque nós não utilizô mais ela, ela ficô abandonada aí, por mode utilizá, tem que gastá um pôco pa podê botá ela no ponto de iniciá, mas se *nós* tivesse zelado, *dêxado* tudo, né, mas confiô, dêxamo abandonada.

O *gênero* do informante foi a sexta variável selecionada como relevante para o uso do pronome *nós*.

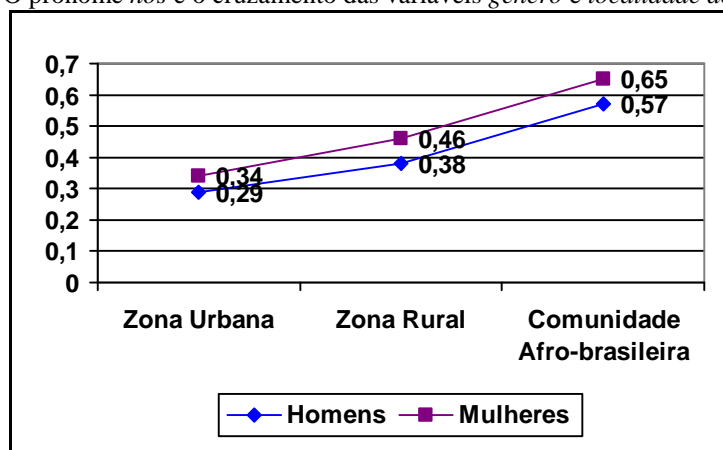
Tabela 28: O pronome *nós* a variável social *gênero do informante*.

Gênero do informante	N / T	%	P. R.
Feminino	169 / 1820	9	.55
Masculino	103 / 1474	7	.44
TOTAL	272 / 3294	8	

Nível de significância .010

Apesar das freqüências se mostrarem praticamente neutras quanto ao uso do pronome *nós* por falantes dos gêneros feminino e masculino, os pesos relativos alcançados por essas duas variantes foram , **55** e ,**44** respectivamente. Esse índices indicam que a estratégia indeterminadora *nós* é favorecida na fala feminina.

Assim como foi feito o cruzamento entre as variáveis sociais *localidade* e *gênero* do informante, a fim de que pudesse ser verificado o processo de implementação do pronome *a gente* no português popular do interior do estado da Bahia; o mesmo foi feito com relação ao pronome *nós*. Os pesos relativos alcançados por cada um dos fatores originados do cruzamento dessas variáveis são apresentados no gráfico 3.

Gráfico 2: O pronome *nós* e o cruzamento das variáveis *gênero* e *localidade do informante*.

Nível de significância .000

Nos três tipos de comunidades estudadas, a fala das mulheres assume a liderança quanto ao favorecimento da estratégia *nós*. Na sede e na zona rural da cidade de Santo Antonio, tanto homens quanto mulheres alcançam pesos relativos abaixo de ,50, que desfavorecem a variante dependente. Nas comunidades afro-brasileiras, ambos os gêneros favorecem o uso do *nós*, contudo as mulheres lideram esse processo: o peso relativo que alcançam é ,**65** contra ,**57** dos homens. Portanto, nos três tipos de comunidade observados, as mulheres exi-

bem um comportamento mais conservador, e os homens, que geralmente têm mais contato com o mundo exterior, lideram a mudança de substituição do *nós* pelo *a gente*.

Os resultados da variável social *escolaridade*, o último fator selecionado nessa rodada, assim como o visto na variável *gênero*, também são discretos quanto à frequência do pronome *nós*. Os fatores que compõem esta variável social demonstram sua relevância quando analisados os seus pesos relativos.

Tabela 29: O pronome *nós* segundo a variável social *escolaridade*.

Escolaridade	N / T	%	P. R.
Semi-analfabeto	146 / 1952	7	.54
Analfabeto	126 / 1342	9	.44
TOTAL	272 / 3294	8	

Nível de significância .010

As frequências praticamente idênticas de 7% e 9% demonstram resultados neutros quanto ao favorecimento do pronome *nós*. Os pesos relativos obtidos, entretanto, indicam o favorecimento do pronome por falantes semi-analfabetos (**.54**) e o desfavorecimento da estratégia na fala de informantes com nenhuma escolarização, em que o peso relativo é de **.44**. Pode parecer contraditório que os falantes com alguma escolarização usem menos a variante padrão que os analfabetos. Entretanto, a escolarização deve ser vista aqui como um índice de mais contato com o mundo exterior. Como a substituição do *nós* pelo *a gente* é um processo de mudança em curso desde os grandes centros urbanos, e natural que os falantes com alguma escolarização usem mais o *a gente*.

4.2.3 Você

Dentre os cinco fatores selecionados pelo programa *VARBRUL* como favorecedores da escolha pelo pronome *ocê* está, novamente, em primeiro lugar, a variável *forma antecedente*. Os cinco fatores foram distribuídos pela seguinte ordem, apresentada na tabela 30.

Tabela 30: Variáveis favorecedoras da estratégia *você*

Variável	
1º	Forma antecedente
2º	Modo verbal
3º	Inclusão do falante
4º	Localidade do falante
5º	Escolaridade

Assim como o que foi feito com relação à análise das variantes *a gente* e *nós*, foram mantidos apenas dois fatores do grupo de fatores que compõe a variável linguística *forma antecedente*: a *primeira menção* e o *você*. As demais variáveis, por suas baixas ocorrências, foram amalgamadas.

Os resultados da variável *forma antecedente* constam na tabela a seguir.

Tabela 31: O pronome *você* segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	221 / 2513	9	.59
Você	126 / 135	93	.99
Outras formas	22 / 1127	2	.19
TOTAL	369 / 3375	10	

Nível de significância .046

Também quanto à variável dependente *você*, confirma-se o princípio do *paralelismo formal* no nível discursivo. Das 135 ocorrências antecederas pelo pronome *você*, 93% são do próprio pronome – o peso relativo desse fator (**,99**) corrobora este resultado, afirmando ser praticamente categórica essa repetição. A *primeira menção* também foi apontada pelo peso relativo de **,59** como favorecedora da estratégia *você*. As demais formas antecederam apenas 2% das ocorrências do *você* e obtiveram **,19** de peso relativo, o que indica desfavorecimento da estratégia.

Quanto ao modo verbal, as ocorrências do pronome *você* são favorecidas pelo modo *Irrealis*, como pode ser visto na tabela 32.

Tabela 32: O pronome *você* segundo a variável *modo verbal*.

Aspecto	N / T	%	P. R.
Realis	112 / 2488	5	.36
Irrealis	257 / 1287	20	.76
TOTAL	369 / 3375	10	

Nível de significância .046

As ocorrências em modo verbal *realis* – como a apresentada no exemplo 123 – atingem peso relativo de ,**36**. No modo *irrealis* (cf. exemplo 124), o peso relativo é de ,**76**.

(123) Lugá ótimo pra você criá filho, né, porque aqui *você* num corre risco de risco de carro.

(124) O posto que tem lá, passa remédio hoje, se o remédio não tivé lá hoje, *você* fô pegá o remédio, num tivé, o médico passa o remédio num tivé naquele posto, se *você* fô amanhã, ela vai falá “A receita foi de ontem.” É assim.

A 3^a variável selecionada pelo programa *VARBRUL* foi a *inclusão do falante no universo de referência do sujeito indeterminado*.

Tabela 33: O pronome *você* segundo a variável *inclusão do falante*.

Inclusão do falante	N / T	%	P. R.
Incluso	365 / 3252	11	.57
Não-incluso	4 / 523	1	.13
TOTAL	369 / 3375	10	

Nível de significância .046

Apesar de observadas apenas quatro ocorrências do pronome *você*, está variável foi selecionada como estatisticamente relevante pelo *VARBRUL*, e o contexto em que a referência não inclui o falante recebeu o peso relativo de ,**13**, demonstrando que raramente esta forma não inclui o próprio falante, quando usada como estratégia de indeterminação do sujeito. As quatro ocorrências são as mesmas apresentadas no exemplo (120) deste capítulo, reapresentada a seguir como (125), no qual o falante utiliza-se do pronome *vocês* para referir-se a um grupo específico de pessoas.

(125) ...a gruta que tinha, **cês** dismantelô... ôta que tinha, **cês** mantelô, assim ondé que tinha, **cês** mantelô e **fez** uma igreja [...].

A distribuição das ocorrências do pronome *você* pela variável social *localidade do informante* consta na tabela 34.

Tabela 34: O pronome *você* segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Com. Afro-brasileira	211 / 2002	11	.57
Zona Rural	67 / 800	8	.42
Zona Urbana	91 / 973	9	.43
TOTAL	369 / 3375	10	

Nível de significância .046

Os pesos relativos obtidos pela sede e pela zona rural da cidade de Santo Antonio são ,43 e ,42, respectivamente; o pronome *você*, entretanto, só é favorecido nas comunidades afro-brasileiras, que alcançam o índice de ,57.

A *escolaridade do informante* foi a última variável selecionada como relevante para a escolha pela estratégia indeterminadora *você*.

Tabela 35: O pronome *você* segundo a variável social *escolaridade*.

Escolaridade	N / T	%	P. R.
Semi-analfabeto	244 / 2223	11	.53
Analfabeto	125 / 1552	8	.46
TOTAL	369 / 3375	10	

Nível de significância .046

Segundo os dados da tabela 35, o acesso a alguma escolaridade, ainda que mínima, favorece, embora de maneira discreta, a escolha pelo pronome *você* – com peso relativo de ,53. O peso relativo alcançado pelos falantes analfabetos é de ,46.

Os resultados dessas variáveis sociais são, em princípio, contraditórios, pois as comunidades rurais afro-brasileiras isoladas têm apresentado o comportamento lingüístico mais distante do padrão e mais conservador. Portanto, é estranho que a variante mais usada pelos afro-brasileiros seja a mesma que é mais freqüente na fala dos que tem alguma escolarização.

4.2.4 Eles

Quanto ao pronome *eles*, foram quatro as variáveis selecionadas pelo programa VAR-BRUL, três variáveis lingüísticas e somente uma variável social, apresentadas por ordem de relevância na tabela 36.

Tabela 36: Variáveis favorecedoras da estratégia *eles*

Variável	
1º	Inclusão do falante
2º	Forma antecedente
3º	Localidade do informante
4º	Modo Verbal

A primeira variável favorecedora do pronome *eles* é a *inclusão do falante no universo de referência do sujeito indeterminado*, cujos resultados são apresentados a seguir.

Tabela 37: O pronome *eles* segundo a variável *inclusão do falante*.

Inclusão do falante	N / T	%	P. R.
Incluso	4 / 2999	-	.26
Não-incluso	217 / 516	42	1.00
TOTAL	221 / 3515	6	

Nível de significância .008

Os dados mostram que, ao utilizar a estratégia *eles*, a não-inclusão do falante é praticamente categórica. O peso relativo alcançado por este fator é 1,00 e se deve ao traço semântico do pronome de 3ª pessoa do plural. As quatro ocorrências do pronome *eles* nas quais o falante foi incluído constam no exemplo (116), aqui rerepresentado como exemplo (126). Nesse contexto, o falante, após ser inquirido pelo documentador sobre que remédio é dado aos doentes, refere-se a um hábito comum da comunidade, na qual se inclui em seu discurso.

(126) DOC: Hum... E que remédio é esse que dá assim?

INF: *Eles* dão o tá da 'um moscada... Ø *rala*, Ø *coloca* na água morna, Ø *dá* pra bebê...

A análise da variável *forma antecedente*, a segunda selecionada como estatisticamente relevante pelo programa *VARBRUL*, mostra que a primeira menção favorece o uso do pronome *eles* e alcança o peso relativo de ,57; porém o fator realmente favorecedor do uso do pronome *eles* é o próprio pronome, cujo peso relativo é de 1,00 (cf. tabela 38), ratificando mais uma vez o princípio do paralelismo formal.

Tabela 38: O pronome *eles* segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	168 / 2473	7	.57
Eles	41 / 46	89	1.00
Ø + V3PS	4 / 365	1	.12
Ø + V3PP	4 / 47	9	.20
Outras formas	4 / 365	1	.41
TOTAL	221 / 3515	6	

Nível de significância .008

A variável social *localidade do informante* também foi selecionada como favorecedora da estratégia indeterminadoras *eles*. Seus resultados são apresentados a seguir.

Tabela 39: O pronome *eles* segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Zona Urbana	86 / 893	10	.39
Zona Rural	7 / 753	1	.18
Com. Afro-brasileira	128 / 1869	7	.70
TOTAL	221 / 3515	6	

Nível de significância .008

Apesar da frequência obtida pelo pronome *eles* na sede da cidade de Santo Antonio ter sido maior do que nas demais localidades, o peso relativo de ,70 aponta serem as comunidades afro-brasileiras as favorecedoras da estratégia. A sede de Santo Antonio atinge ,39 e a zona rural dessa cidade, apenas ,18.

O modo verbal foi a quarta variável selecionada na análise do pronome *eles* e os seus resultados são apresentados na tabela 40.

Tabela 40: O pronome *eles* segundo a variável *modo verbal*.

Aspecto	N / T	%	P. R.
Realis	171 / 2367	7	.45
Irrealis	50 / 1148	4	.60
TOTAL	221 / 3515	6	

Nível de significância .008

Verifica-se que, apesar da frequência alcançada pelo modo *realis* ser superior à frequência do modo *irrealis*, este último fator é que favorece o uso do pronome *eles*, pois atinge peso relativo de ,60, enquanto o modo *realis* atinge ,45. No exemplo (127) constam duas ocorrências do pronome *eles*: a primeira refere-se a um fato real e é classificada como *realis*, a segunda, em que o falante formula uma hipótese, é classificada como *irrealis*.

(127) É *eles* num dão oportunidade pas pessoa, se *eles* pegasse lá, os menino... lá muitas mãe faz de tudo pos menino ir po colégio.

4.2.5 Ø + V3PP

Na análise quantitativa dos dados referentes à forma indeterminadora Ø + V3PP, uma variante padrão, o programa VARBRUL selecionou como favorecedoras desta estratégia sete variáveis, apresentadas por ordem de relevância na tabela a seguir.

Tabela 41: Variáveis favorecedoras da estratégia Ø + V3PP

Variável	
1°	Nível de referencialidade
2°	Forma antecedente
3°	Localidade do falante
4°	Estada fora da comunidade
5°	Gênero
6°	Tipo de frase
7°	Modo Verbal

Os resultados da primeira variável selecionada, o *nível de referencialidade do sujeito indeterminado*, constam na tabela 42.

Tabela 42: A forma $\emptyset + V3PP$ segundo a variável *Nível de referencialidade*.

Nível de referencialidade	N / T	%	P. R.
Genérico	3 / 1472	-	.37
Parcialmente definido	3 / 1103	-	.27
Específico indefinido	154 / 440	35	.98
TOTAL	160 / 3045	5	

Nível de significância .013

Na tabela, observa-se a baixa ocorrência da forma indeterminadora $\emptyset + V3PP$ nos níveis de referencialidade *genérico* e *grupo parcialmente definido*, que conseguem apenas três ocorrências cada. O nível em que prevalece a forma $\emptyset + V3PP$ é o do *agente específico indefinido*, nível em que 35% das ocorrências são dessa forma. O peso relativo desse fator é de ,98, indicando que o uso da terceira pessoa do plural como estratégia de indeterminação se relaciona de maneira praticamente categórica com a referência a um agente específico indefinido. O uso da forma *eles* nos outros níveis de referencialidade seria fortemente marginal.

A variável *forma antecedente* apresentou os seguintes resultados:

Tabela 43: A forma $\emptyset + V3PP$ segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	117 / 1996	6	.60
$\emptyset + V3PP$	35 / 47	74	.90
Outras formas	8 / 1002	1	.29
TOTAL	160 / 3045	5	

Nível de significância .013

O princípio do paralelismo formal aplica-se mais uma vez, pois o uso da forma $\emptyset + V3PP$ passa de 05% para 74% quando essa mesma forma é usada na oração anterior para indeterminar o sujeito, com peso relativo de ,90.

A variável social *localidade do informante* foi a terceira selecionada pelo programa *VARBRUL*. Os pesos relativos obtidos pela sede e pela zona rural da cidade de Santo Antonio de Jesus favorecem o uso da forma $\emptyset + V3PP$, desfavorecida nas comunidades afro-brasileiras, que obtém ,35 de peso relativo, enquanto a sede alcança ,69 e a zona rural, ,67.

Tabela 44: A forma $\emptyset + V3PP$ segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Com. Afro-brasileira	67 / 1675	4	.35
Zona Rural	22 / 668	3	.67
Zona Urbana	71 / 702	10	.69
TOTAL	160 / 3045	5	

Nível de significância .013

Os resultados da variável social *estada fora da comunidade*, apresentados na tabela 45, também se mostraram relevantes para a escolha pela forma indeterminadora $\emptyset + V3PP$.

Tabela 45: A forma $\emptyset + V3PP$ segundo a variável social *estada fora da comunidade*.

Estada fora da comunidade	N / T	%	P. R.
Sim	79 / 1317	6	.58
Não	81 / 1649	5	.43
TOTAL	160 / 3045	5	

Nível de significância .013

Conforme o esperado do comportamento de uma estratégia padrão, a forma $\emptyset + V3PP$ é favorecida por falantes que saíram da comunidade por um período superior a seis meses e desfavorecida pelos falantes que não saíram. O peso relativo desse último fator é de ,43, enquanto o peso do fator favorecedor da estratégia indeterminadora $\emptyset + V3PP$ é ,58.

No tocante à variável *gênero do informante*, enquanto o número de ocorrências e as frequências são similares entre os dois gêneros, os pesos relativos se distanciam em 14 pontos: as mulheres alcançam ,44 e os homens ,58. Esses dados mostram serem os homens favorecedores da estratégia padrão $\emptyset + V3PP$.

Tabela 46: A forma $\emptyset + V3PP$ segundo a variável social *gênero do informante*.

Gênero do informante	N / T	%	P. R.
Feminino	80 / 1737	5	.44
Masculino	80 / 1308	6	.58
TOTAL	160 / 3045	5	

Nível de significância .013

Na análise da variável *tipo de frase* não foram encontradas ocorrências dessa variante em orações interrogativas. Quanto às orações declarativas e negativas, a distribuição das o-

corrências, apresentada na tabela 47, aponta um discreto favorecimento por parte das orações declarativas, cujo o peso relativo é ,53; as orações negativas em que figuram a construção \emptyset + V3PP são apenas oito e alcançam peso relativo de ,25, revelando-se um contexto desfavorável dessa estratégia.

Tabela 47: A forma \emptyset + V3PP segundo a variável *tipo de frase*.

Tipo de Frase	N / T	%	P. R.
Declarativa	152 / 24731	6	.53
Negativa	8 / 314	3	.25
TOTAL	160 / 3045	5	

Nível de significância .013

A última variável selecionada pelo programa VARBRUL foi o *modo verbal*. Seus resultados são apresentados a seguir, na tabela 48.

Tabela 48: A forma \emptyset + V3PP segundo a variável *modo verbal*.

Aspecto	N / T	%	P. R.
Realis	154 / 1964	8	.60
Irrealis	6 / 1081	1	.32
TOTAL	160 / 3045	5	

Nível de significância .013

As ocorrências em que o modo verbal é classificado como *realis* favorecem com peso relativo de ,60 o uso da variante dependente \emptyset + V3PP; o modo verbal *irrealis* desfavorece o uso da mesma, com peso relativo de ,32. É natural a correlação entre a forma que é utilizada normalmente com referência a agentes específicos indefinidos e a narração de fatos objetivos. As orações em (128) e (129) exemplificam as ocorrências de \emptyset + V3PP nos citados modos verbais. No exemplo (128), o falante se utiliza da forma \emptyset + V3PP para tratar de um ato real já executado; já o exemplo (129) apresenta uma hipótese levantada pelo informante.

(128) Tinha muito mato, já derrubar.

(129) Vai vê que já [criar] estrada pra lá...

4.2.6 Ø + V3PS

O pacote de programas *VARBRUL*, no tratamento quantitativo dos dados referentes à forma Ø + V3PS, selecionou as seguintes variáveis como relevantes do ponto de vista estatístico:

Tabela 49: Variáveis favorecedoras da estratégia Ø + V3PS

Variável	
1º	Forma antecedente
2º	Nível de referencialidade
3º	Escolaridade
4º	Estada fora da comunidade
5º	Tipo de verbo
6º	Localidade do falante

Segundo dados provenientes da análise quantitativa realizada, também com relação à forma Ø + V3PS pode ser constatado o paralelismo formal, conforme pode ser visto na tabela 50.

Tabela 50: A forma Ø + V3PS segundo a variável *forma antecedente*.

Forma antecedente	N / T	%	P. R.
1ª menção	612 / 2513	24	.56
Ø + V3PS	318 / 368	86	.96
Outras formas	33 / 894	4	.12
TOTAL	963 / 3775	26	

Nível de significância .008

As ocorrências de Ø + V3PS antecedidas por esta mesma forma têm peso relativo de **,96**. A *primeira menção* também se mostrou favorecedora da estratégia: o peso relativo deste fator é **,56**.

Quanto ao *nível de referencialidade*, a forma Ø + V3PS foi a única a distribuir-se pelos quatro níveis de referencialidade estudados nesta pesquisa. Os dados dispostos na tabela 51 mostram que a maior frequência da estratégia Ø + V3PS (38%), assim como o maior peso relativo (**,67**) foram obtidos no nível de referencialidade *genérico*. Em seguida, observamos o leve favorecimento da estratégia, através do peso relativo de **,55**, no nível de referencialidade

do *agente específico indefinido*. Os dois níveis restantes, o do *grupo parcialmente definido* e do *eu ampliado* desfavorecem o uso da forma $\emptyset + V3PS$, pois obtém, respectivamente, índices de ,39 e ,24.

Tabela 51: A forma $\emptyset + V3PS$ segundo a variável *nível de referencialidade*.

Nível de referencialidade	N / T	%	P. R.
Genérico	582 / 1547	38	.67
Parcialmente definido	202 / 1204	17	.39
Eu ampliado	41 / 543	8	.24
Específico indefinido	138 / 481	29	.55
TOTAL	963 / 3775	26	

Nível de significância .008

A terceira variável selecionada como favorecedora da estratégia $\emptyset + V3PS$ foi a primeira variável social, a *escolaridade do informante*, cujos resultados são apresentados na tabela a seguir.

Tabela 52: A forma $\emptyset + V3PS$ segundo a variável social *escolaridade*.

Escolaridade	N / T	%	P. R.
Semi-analfabeto	395 / 2223	18	.43
Analfabeto	568 / 1552	37	.60
TOTAL	963 / 3775	26	

Nível de significância .008

Conforme esperávamos constatar, a forma $\emptyset + V3PS$ é mais freqüente entre os falantes sem nenhuma escolaridade: chega a ser o dobro da freqüência da estratégia em questão na fala dos semi-alfabetizados. A análise dos pesos relativos também demonstra o favorecimento da estratégia indeterminadora $\emptyset + V3PS$ por parte dos analfabetos – o peso relativo dessa variante é ,60.

Quanto à segunda variável social selecionada, a *estada fora da comunidade*, os resultados também foram os esperados levadas em conta as hipóteses que balizam este estudo.

Tabela 53: A forma $\emptyset + V3PS$ segundo a variável social *estada fora da comunidade*.

Estada fora da comunidade	N / T	%	P. R.
Sim	357 / 1734	21	.45
Não	606 / 2041	30	.54
TOTAL	963 / 3775	26	

Nível de significância .008

Informantes que saíram das suas comunidades por um período mínimo de seis meses alcançam peso relativos de **,45**, o que desfavorece o uso da estratégia $\emptyset + V3PS$. O peso relativo obtido por falantes que não saíram de suas comunidades é **,54**, que aponta favorecimento, mesmo que de forma discreta, da utilização da forma $\emptyset + V3PS$.

A análise dos pesos relativos dos *tipos verbais* observados neste estudo pode ser feita a partir dos dados da tabela 54.

Tabela 54: A forma $\emptyset + V3PS$ segundo a variável *tipo de verbo*.

Tipo de verbo	N / T	%	P. R.
Transitivos de ação pontual	455 / 1432	32	.52
Transitivos de ação não pontual	201 / 813	25	.59
Cognitivos	18 / 62	22	.42
Discendi	52 / 105	50	.80
Intransitivos	41 / 169	24	.43
Intransitivos inacusativos	18 / 60	30	.74
Movimento / Locativos	66 / 402	16	.34
Estativos/ligação	28 / 246	11	.28
Verbos suporte/leve	8 / 55	15	.39
Posse	9 / 97	9	.25
Existenciais	7 / 60	12	.36
Causativos	56 / 228	25	.62
Sem carga semântica	4 / 21	19	.41
TOTAL	963 / 3775	26	

Nível de significância .008

Apenas cinco tipos de verbo, segundo a nossa classificação, se mostraram favorecedores da estratégia $\emptyset + V3PS$: verbos transitivos de ação pontual e não pontual, discendi, intransitivos inacusativos e causativos. Dentre estes, o que mais favorece a forma $\emptyset + V3PS$

são verbos classificados como *discendi*, que alcançam peso relativo de ,80. Isso se deve, principalmente, a estruturas como a apresentada no exemplo (130), em que o falante se utiliza da construção *diz que*. Verbos intransitivos inacusativos, no exemplo (131), aparecem em segundo lugar quanto ao favorecimento da estratégia, com peso relativo de ,74 e são seguidos pelos verbos causativos (132), com ,62. Os verbos transitivos de ação pontual (133) e não pontual (134) apresentam pesos relativos de ,52 e ,59.

(130) DOC: Você já ouviu falá de algum... gente que foi atacado aí por bicho, aí nesse mar?

INF: 'Inda não. Só de vez em quando, que \emptyset diz que parece... parece baleia, assim, na bêra da praia...

(131) Aqui, quando \emptyset morre, o povo fica até de manhã lá.

(132) Depois... e ... e forra com... com a toalha e coloca os prato, só as verdura, né, e a carne de frango e os copo de suco e bala de junto, né. Aí \emptyset manda as criança se sentá.

(133) É... tem esses benefício mesmo do PETE. Esse salário, se num for po colégio, \emptyset corta.

(134) E esse ano é a mesma coisa, agora num \emptyset sabe se assim vai torná continuá a mesma coisa.

A última variável selecionada pelo programa *VARBRUL* é a localidade do informante, em que, mais uma vez, os resultados corroboram com as hipóteses defendidas aqui.

Tabela 55: A forma \emptyset + V3PS segundo a variável social *localidade do informante*.

Localidade do informante	N / T	%	P. R.
Com. Afro-brasileira	605 / 2002	30	.55
Zona Rural	196 / 800	24	.47
Zona Urbana	162 / 936	17	.43
TOTAL	963 / 3775	26	

Nível de significância .008

Não só a frequência da forma $\emptyset + V3PS$ se eleva a medida que há aproximação das comunidades afro-brasileiras, os pesos relativos também refletem um contínuo lingüístico no que diz respeito à utilização da variável $\emptyset + V3PS$ nas localidades estudadas. Enquanto, na sede do município de Santo Antonio de Jesus, a frequência da forma $\emptyset + V3PS$ é de 17% e o peso relativo é ,43 (índices que indicam o desfavorecimento desta variável dependente), na zona rural da mesma cidade a frequência desta foram sobe para 24% e o peso relativo alcança ,47. O favorecimento, contudo, somente é observado no vernáculo das comunidades afro-brasileiras: a estratégia indeterminadora $\emptyset + V3PS$ alcança 30% de uso e peso relativo de ,55.

Todas as variáveis sociais selecionadas pelo VARBRUL como estatisticamente relevantes confirmam a hipótese de que a estratégia $\emptyset + V3PS$, resultante de um processo de erosão da morfologia aparente do português, estaria correlacionada aos cenários sociolingüísticos mais diretamente ligados ao contato entre línguas na sua formação. Assim, o maior índice da variante $\emptyset + V3PS$ encontra-se nas comunidades afro-brasileiras, cuja formação está mais diretamente ligada ao contato entre línguas. Quanto ao vetor contrário, a influência do padrão lingüístico dos grandes centros urbanos, observa-se ser predominante na zona urbana do município – nessa localidade se encontra o menor índice de emprego da variante $\emptyset + V3PS$. Os resultados das variáveis *estada fora da comunidade* e *nível de escolaridade* representam a evidência empírica de que a substituição da estratégia mais desviante do padrão constitui uma mudança de cima para baixo, dentro da concepção laboviana, decorrente de influências externas às comunidades estudadas. Os falantes com alguma escolaridade e que viveram fora das comunidades, que, por suposto, assimilam mais as influências externas e têm mais contato com o padrão lingüístico urbano, são aqueles que menos usam a variante $\emptyset + V3PS$.

4.3 O SE

Quanto à partícula *se*, foram verificadas nas comunidades afro-brasileiras um total de dez ocorrências. Dessas ocorrências duas, apresentadas nos exemplos (135) e (136), podem ser classificadas como *frases feitas* ou *expressões fixas* e, em uma, apresentada no exemplo (137), há a utilização da partícula *se* a partir da influência do documentador.

(135) Jequié, cidade linda de *se* vê!

(136) Aí chegaro esses pessoal que tem mais, *como se diz*, mais podê no Mêi.

(137) DOC: *Se* usa muito?

INF: *Se* usa muito tê filho em casa.

As ocorrências foram encontradas na fala de dez informantes das comunidades afro-brasileiras, dentre eles seis homens e quatro mulheres. Não foi encontrada mais do que uma ocorrência na fala de cada um desses informantes. Esse baixo número de ocorrências revela que o *se* indeterminador está praticamente ausente da gramática dessas comunidades, predominado a variante $\emptyset + V3PS$, desprovida de uma marca morfológica explícita de indeterminação.

Já na sede e na zona rural da cidade de Santo Antonio de Jesus foram contabilizadas 46 ocorrências da partícula *se*. Chegaram a ser encontradas até onze ocorrências na fala de um único informante. Assim como nas comunidades afro-brasileiras, também foi verificada a ocorrência da expressão fixa *como se diz* (cf. exemplo 138), contudo em maior número: foram encontradas seis dessas ocorrências.

(138) Aquela casa é de... é ...como é que **se** diz, é dos associados daqui da zona rural que se associa ali.

Apenas em três ocorrências (cf. exemplos 139-141) foram verificados objetos no plural, porém em nenhuma dessas houve concordância com o argumento interno.

(139) O home que mora na zona rural, na zona do campo, num pode **se** incruzá *os braço*, num pode ININT, num pode incruzá *os braço*, né?

(140) É, poque uma... uma bomba de vinte litro **se** bota *6 ml*.

(141) Não **se** pode prantá *certos tipo de coisa*.

Ao se comparar o número de ocorrências obtido na zona rural e na sede do município de Santo Antonio de Jesus com o obtido nas comunidades afro-brasileiras, fica claro que naquelas percebe-se a maior influência dos grandes centros de irradiação lingüística do país, o

que permite aos seus falantes a assimilação mais intensa da variante padrão. Contudo o emprego da partícula *se* como índice de indeterminação do sujeito ainda é muito baixo nas variedades populares o interior do país.

5 CONSIDERAÇÕES FINAIS

Na introdução desse trabalho, foram estabelecidos três objetivos a serem alcançados: (i) descrever as estratégias de indeterminação do sujeito no português popular do interior do estado da Bahia; (ii) verificar o encaixamento da variação na estrutura lingüística e social das comunidades estudadas; (iii) contrastar o comportamento lingüístico, quanto à indeterminação do sujeito, dos falantes com pouca ou nenhuma escolarização de comunidades afro-brasileiras isoladas e do português popular da cidade de Santo Antonio de Jesus.

Ao serem descritas as estratégias que compõem o quadro de variação concernente à indeterminação do sujeito pronominal no português popular do interior do estado da Bahia, observou-se um elenco bem maior de estratégias do que prescreve a tradição gramatical: no total, foram encontradas dez estratégias de indeterminação, três verbais e sete pronominais, cujas ocorrências são foneticamente realizadas em setenta por cento dos casos.

Foram descritas e analisadas ocorrências dos pronomes indeterminadores *nós*, *a gente*, *você* e *eles* e das formas verbais $\emptyset + V3PP$ e $\emptyset + V3PS$. Dentre essas, a mais utilizada é a estratégia *a gente*, responsável por 47% das ocorrências de indeterminação do sujeito, a segunda estratégia em número de ocorrências é a forma $\emptyset + V3PS$, que alcança freqüência de 26%, seguida pelas formas *você*, *nós*, *eles* e $\emptyset + V3PP$, com, respectivamente, 10%, 7%, 6% e 4% de uso.

Assim, foi verificada a distribuição das seis estratégias de indeterminação do sujeito contempladas nesse estudo pelos quatro níveis de referencialidade aqui estudados, e constatou-se que há três estratégias principais para cada um desses níveis. No nível genérico, em que a distribuição das ocorrências se deu de modo mais homogêneo, três formas distintas prevalecem: *a gente*, *você* e $\emptyset + V3PS$. Nos níveis *grupo parcialmente definido* e *eu ampliado* são encontrados predominantemente os pronomes de primeira pessoa do plural, *a gente* e *nós*, e a forma $\emptyset + V3PS$. Quanto ao nível do *agente específico indefinido*, o predomínio é das formas de terceira pessoa do plural, *eles* e $\emptyset + V3PP$, contudo a forma $\emptyset + V3PS$ possui boa representatividade nesse nível, em que é responsável por um terço das ocorrências. A estratégia $\emptyset + V3PS$ foi a única a se distribuir pelos quatro níveis de referencialidade do sujeito indeterminado, demonstrando uma alta versatilidade.

Quanto aos contextos lingüísticos favorecedores do uso das seis formas indeterminadoras analisadas, identificamos como o principal a *forma antecedente*, única variável selecionada pelas seis estratégias de indeterminação analisadas, e sempre em primeiro ou segundo

lugar. Desse modo, confirmou-se o princípio do *paralelismo formal* segundo o qual, após a escolha de uma forma indeterminadora, esta tende a se repetir.

A variável *nível de referencialidade* favorece a escolha de quatro das seis estratégias indeterminadoras – os pronomes de primeira pessoa do plural e as formas $\emptyset + V3PP$ e $\emptyset + V3PP$ – e, assim como a variável forma antecedente, quando selecionada, ocupou a primeira ou segunda posição dentre as estratégias favorecedoras. Mais uma vez, foi observada a distribuição das estratégias pelos níveis de referencialidade segundo a carga semântica e a capacidade de generalização: enquanto os pronomes *a gente* e *nós* são favorecidos sobretudo pelo nível do *eu ampliado*, a forma indeterminadora $\emptyset + V3PP$ se relaciona de maneira praticamente categórica com a referência a um agente específico indefinido e a estratégia $\emptyset + V3PS$ é favorecida principalmente pelo nível genérico.

O modo verbal foi um fator selecionado por quatro das variantes dependentes estudadas. O pronome *nós* e a forma $\emptyset + V3PP$ foram favorecidos pelo modo *realis*, ou seja, nas ocorrências que se refere a um fato real. O modo *irrealis* favorece as estratégias *você* e *eles*. Também para estes dois últimos pronomes, a *inclusão do falante* no universo de referência do sujeito indeterminado é relevante – o pronome *você*, quando usado como estratégia de indeterminação do sujeito, raramente exclui o próprio falante. Com relação ao pronome *eles*, a não-inclusão do falante é praticamente categórica.

Não se mostrou muito relevante quanto à escolha das estratégias indeterminadoras a influência das variáveis *tipo de frase* e *tipo de verbo* – cada uma dessas variáveis só foi selecionada por uma forma de indeterminação do sujeito. Quanto ao tipo de frase, as orações declarativas favorecem discretamente o uso da forma $\emptyset + V3PP$; quanto ao tipo de verbo, os verbos transitivos de ação pontual e não pontual, os intransitivos inacusativos, os causativos e, sobretudo, os discendi – classificação à qual pertence a construção *diz que* – são os que mais favorecem a forma $\emptyset + V3PS$.

Quanto aos contextos sociais que afetaram a escolha das formas de indeterminação do sujeito, destaca-se a influência da variável *localidade do informante*, a única variável social a ser selecionada como estatisticamente relevante pelas seis estratégias de indeterminação do sujeito. A existência de um *continuum* lingüístico, no que concerne ao sujeito indeterminado, proveniente dos processos de transmissão irregular da língua, é atestada pelas frequências de uso e pelos pesos relativos da forma $\emptyset + V3PS$ nas localidades analisadas neste trabalho. A categoria vazia apresenta maior funcionalidade nas comunidades afro-brasileiras do que nas

demais localidades estudadas por se manterem nelas os traços da aquisição defectiva do português.

Estada fora da comunidade se mostrou relevante quanto à escolha pelas formas *a gente*, $\emptyset + V3PP$ e $\emptyset + V3PS$, e os resultados corroboram com a hipótese de que difusão do português padrão a partir grandes centros urbanos. O pronome *a gente* e a estratégia padrão $\emptyset + V3PP$ são favorecidos pelos falantes que saíram da comunidade, enquanto a forma $\emptyset + V3PS$, uma variante não padrão, pelos que não tiveram contato com o exterior. No que concerne à escolaridade, essa variável, juntamente com a estada fora da comunidade, pode ser vista aqui como um índice de mais contato com o mundo exterior. Dessa forma, os resultados encontrados foram os esperados: falantes quem já tiveram contato com a escola favorecem o uso do pronome *nós*, enquanto a forma $\emptyset + V3PS$, não padrão, é mais freqüente entre os falantes sem nenhuma escolaridade.

O caráter lingüístico conservador da fala feminina é atestado, no que diz respeito à indeterminação do sujeito, pelo favorecimento da estratégia *nós*. Os dados da análise da forma $\emptyset + V3PP$ mostraram serem os homens favorecedores dessa estratégia.

A variável faixa etária se mostrou importante principalmente quanto às conclusões a cerca dos pronomes *a gente* e *nós*: a substituição do pronome *nós* pelo *a gente* é verificada não só quando o sujeito é referencial, mas também nos contextos de indeterminação.

A fim de que se pudesse verificar o processo de implementação do pronome *a gente* no português popular do interior do estado da Bahia, foi realizado, na análise dessa estratégia indeterminadora, o cruzamento entre as variáveis *faixa etária* e *localidade do informante*. A zona urbana lidera a mudança em favor da forma inovadora; e o uso dessa estratégia é crescente na medida em que se passa para as gerações mais novas. O cruzamento das variáveis sociais *gênero* e *localidade do informante*, verificado na análise da estratégia *nós*, releva que, nos três tipos de comunidades estudadas, as mulheres exibem um comportamento mais conservador, e os homens, que geralmente têm mais contato com o mundo exterior, lideram a mudança de substituição do *nós* pelo *a gente*. Há, portanto, resultados empíricos que fundamentariam uma generalização no sentido da implementação do *a gente* como estratégia de indeterminação no português popular do interior do Estado da Bahia. A substituição seria liderada pelos núcleos urbanos do interior do Estado e, em seguida se difundiria pela zona rural, atingindo, posteriormente, as comunidades rurais mais isoladas.

Devido ao número reduzido de ocorrências da partícula *se*, a análise do seu comportamento do português popular se deu de forma qualitativa. Quando comparados o número de

ocorrências obtido na zona rural e na sede do município de Santo Antonio de Jesus com o obtido pelas comunidades afro-brasileiras observou-se que tanto a zona rural quanto a urbana recebem mais influência dos grandes centros de irradiação lingüística do país, assimilando um pouco mais a variante padrão. De qualquer forma, esta pesquisa constatou que o uso da variante padrão para a indeterminação do sujeito é extremamente reduzido na fala popular do interior do Estado da Bahia.

O conjunto dos resultados obtidos e, sobretudo, o comportamento da estratégia $\emptyset + V3PS$ permite-nos confirmar as hipóteses levantadas ao longo da dissertação, fundamentadas pelo conceito da *transmissão lingüística irregular* e da realidade lingüística bipolar, através dos quais são explicados os aspectos morfossintáticos que distinguem e, ao mesmo tempo, aproximam os dialetos rurais da norma culta. A aproximação entre o comportamento lingüístico desses dois pólos se dá pela dissolução das marcas mais fortes da aquisição imperfeita do PB na fala das camadas populares da sociedade e pelo afrouxamento dos padrões lingüístico das camadas médias e altas, que se inicia com os processos de independência do país. Por causa do nivelamento lingüístico que atinge as variedades populares do PB, propiciado pela recente e intensa urbanização do país e pela vulgarização do sistema de ensino e da mídia, as características lingüísticas crioulizantes só são encontradas em poucas comunidades rurais afro-brasileiras isoladas, como as estudadas aqui.

Acredita-se ser este trabalho apenas um dos passos em direção de um aprofundamento acerca da indeterminação do sujeito no português popular do estado da Bahia, a que se pretende dar continuidade através de estudos que, além de comunidades rurais, contemplem também a norma lingüística dos grandes centros urbanos, como a cidade de Salvador.

REFERÊNCIAS

- ARENDS, Jacques; MUYSKEN, Pieter; SMITH, Norval (1995). *Pidgind and Creoles, an introduction*. Philadelphia: John Benjamins.
- ANTONINO, Vívian (2007). *A concordância nominal em predicativos do sujeito e em estruturas passivas no português popular do interior do estado da Bahia*. Dissertação de Mestrado. Instituto de Letras, UFBA, Salvador.
- BAPTISTA, Marlyse (1995). On the Nature of Pro-drop in Capeverdean Creole. In: EPSTEIN, S.; THRÁINSSON, H.; PETER, S. (ed.). *Harvard Working Papers in Linguistics* 5. p. 3-18.
- BAXTER, Alan; LUCCHESI, Dante (1995). A variação na concordância verbal com a 1ª pessoa do singular como parte do processo de descrioulização do dialeto de Helvécia – Ba. *47ª Reunião Nacional da SBPC*, São Luis – Ma.
- BAXTER, Alan; LUCCHESI, Dante (1997). A relevância dos processos de pidginização e criouliização na formação da língua portuguesa no Brasil. *Estudos Lingüísticos e Literários*. Salvador, n. 19, p. 65-84.
- BAXTER, Alan; LUCCHESI, Dante; GUIMARÃES, Maximiliano (1997). "Gender agreement as a decreolizing feature of the Afro-Brazilian rural dialect of Helvécia", *Journal of Pidgin and Creole Language*, 12(1): 1-57.
- BENVENISTE, Émile (1995 [1946]). *O homem na linguagem*. Lisboa: Vega.
- BECHARA, Evanildo (1977). *Moderna gramática portuguesa*. São Paulo: Nacional.
- BECHARA, Evanildo (1992). *Moderna gramática portuguesa: cursos de 1º e 2º graus*. 34 ed. São Paulo: Nacional.
- BORTONI-RICARDO, Stella Maris (2005). *Nós chegemu na escola, e agora?* São Paulo: Parábola.
- CALLOU, Dinah; LEITE, Yonne (2002). *Como falam os brasileiros*. Rio de Janeiro: Jorge Zahar.
- CÂMARA JR., Joaquim Mattoso (1975). *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão.
- CASTILHO, Ataliba de (1992). O português do Brasil. In: ILARI, Rodolfo. *Linguística Românica*. 3. ed. São Paulo.
- CAVALCANTE, Rerisson (2007). *A negação pós-verbal no português brasileiro: análise descritiva e teórica de dialetos rurais de afro-descendentes*. Dissertação de Mestrado. Instituto de Letras, UFBA, Salvador.

- CAVALCANTE, Sílvia Regina de Oliveira (1999). *A indeterminação do sujeito na escrita padrão: a imprensa carioca dos séculos XIX e XX*. Dissertação de Mestrado. Faculdade de Letras, UFRJ, Rio de Janeiro.
- CEGALLA, Domingos Paschoal (1997). *Novíssima gramática da língua portuguesa*. 40 ed. São Paulo: Nacional.
- CHAVES DE MELO, Gladstone (1946). *A Língua do Brasil*. Rio de Janeiro: Agir.
- CHAMBERS, J. K.; TRUDGILL, Peter (1980). *Dialectology*. Cambridge: Cambridge University Press.
- CHOMSKY, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- CUNHA, Cláudia de S. (1993). *Indeterminação Pronominal do Sujeito*. Dissertação de mestrado. Faculdade de Letras, UFRJ, Rio de Janeiro.
- CUNHA, Cláudia de S. (1996). Nós, Você e A Gente: A Influência do Fator Sexo na Hora da Indeterminação. In: LOPES, Célia et al. (org.). *Sexo: uma variável produtiva*. Rio de Janeiro: Faculdade de Letras, UFRJ.
- DUARTE, M. E. L. (1995). *A perda do princípio "evite pronome" no português brasileiro*. Tese de doutorado. Instituto de Estudos da Linguagem, Unicamp, Campinas.
- DUARTE, M. E. L. (2002a). Realizaram, realizou-se ou realizamos. . .? As formas de indeterminação do sujeito em cartas de jornais no século XIX. In: *Para a História do Português Brasileiro - Notícias de Corpora e Outros Estudos*. Rio de Janeiro: in-fólio. v. 4, p. 155-165.
- DUARTE, M. E. L. (2002b). Construções com o se passivador e indeterminador em anúncios do século XIX. In: ALKIMIM, Tânia. *Para a História do Português Brasileiro*. São Paulo: Humanitas, v. 4., p. 155-176.
- DUARTE, M. Eugênia L. (2003). O sujeito expletivo e as construções existenciais. In: RONCARATI, Cláudia (org.). *Português Brasileiro – contato lingüístico, heterogeneidade e história*. Rio de Janeiro: Viveiros de Castro. p. 123-131.
- ELIA, Sílvio (1979). *A unidade lingüística do Brasil: condicionamentos geo-econômicos*. Rio de Janeiro: Padrão.
- FERNÁNDEZ, Francisco Moreno (1998). *Princípios de sociolingüística e sociologia del lenguaje*. Barcelona: Ariel.
- FERREIRA, Carlota et al (1988). Remanescente de um falar crioulo brasileiro (Helvécia-Bahia). In: FERREIRA, Carlota et al. *Diversidade do Português do Brasil - estudos de dialectologia rural e outros*. Salvador: UFBa-PROED. p. 21-32.
- GUY, Gregory (1981). *Linguistic Variation in Brazilian Portuguese: Aspects of Phonology, Syntax and Language History*. PhD Dissertation, University of Pennsylvania, Ann Arbor: University Microfilms International.

GUY, Gregory (1989). On the nature and origins of Popular Brazilian Portuguese. In: *Estudos sobre el Español de América y Lingüística Afroamericana*. Bogotá: Instituto Caro y Cuervo. p. 227-245.

ILARI, Rodolfo *et al* (1996). Os pronomes pessoais do português falado: roteiro para análise. In: BASÍLIO, Margarida; CASTILHO, Ataliba Teixeira (org.). *Gramática do Português Falado: estudos descritivos*. Campinas: Editora da UNICAMP; São Paulo: FAPESP. p. 79-166. (Série Pesquisas, v. IV)

KREUTZ, Lúcio (2000). A educação de imigrantes no Brasil. In: LOPES, Eliana *et alli* (org.). *500 anos de educação no Brasil*. 2 ed. Belo Horizonte: Autêntica. p. 347-370.

LIPSKI, John (1999). Null subjects in Romance-derived creoles: routes of evolution. *Society for Pidgin and Creole Linguistics annual meeting*, Los Angeles.

LOBO, Tânia Conceição Freire (1994). Variantes nacionais do português: sobre a questão da definição do português do Brasil. *Revista Internacional de língua Portuguesa*. Lisboa/AULP, n. 12.

LOBO, Tânia Conceição Freire (2003). A questão da periodização da história lingüística do Brasil. In: CASTRO, Ivo; DUARTE, Inês (org.). *Razões e emoção: miscelânea de estudos em homenagem a Maria Helena Mira Mateus*. Lisboa: Nacional; Casa da Moeda.

LUCCHESI, Dante (1994). Variação e norma: elementos para uma caracterização sócio-lingüística do português do Brasil. *Revista Internacional de Língua Portuguesa*, 12. p. 17-28.

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

LUCCHESI, Dante (2000). *A variação na concordância de gênero em uma comunidade de fala afro-brasileira: novos elementos sobre a formação do português popular do Brasil*. Tese de Doutorado. Faculdade de Letras, UFRJ, Rio de Janeiro.

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

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

LUCCHESI, Dante (2003b). O contato entre línguas e o conhecimento da linguagem humana. *Revista do GELNE*. v.5, n.1 e 2, p.55-62.

MATEUS, Maria Helena Mira *et al* (2003). *Gramática da Língua Portuguesa*. 5 ed. Lisboa: Caminho.

MATTOSO, Kátia (1990). *Ser escravo no Brasil*. 3 ed. São Paulo: Brasiliense.

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

MELO, Gladstone Chaves de (1946). *A Língua do Brasil*, Rio de Janeiro: Agir.

MENDONÇA, Renato (1973 [1933]). *A influência africana no português do Brasil*. 4. ed. Rio de Janeiro: Civilização Brasileira.

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

MIOTO, C; SILVA, M. C. F.; VASCONCELLOS, R. E. (2005). *Novo manual de sintaxe*. 2a ed. Florianópolis: Insular.

NARO, Anthony J. (1976). The gênesis of the reflexive impersonal in Brazilian Portuguese: a study in syntactic changes as a surface phenomenon. *Language*, n.52, p. 779-810.

NARO, Anthony; SCHERRE, Marta (1993a). Sobre as origens do português popular do Brasil, *D.E.L.T.A.*, vol. 9, nº Especial: 437-454.

NARO, Anthony J.; SCHERRE, Maria Marta Pereira (1993b). Duas dimensões do paralelismo formal na concordância verbal no português popular do Brasil. *D.E.L.T.A.*, v.9, n. 1.

NARO, Anthony J.; SCHERRE, Maria Marta Pereira (2003a). Análise quantitativa e tópicos de interpretação do Varbrul. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto. p. 147-177.

NARO, Anthony J.; SCHERRE, Maria Marta Pereira (2003b). O conceito de transmissão linguística irregular e as origens estruturais do português brasileiro: um tema em debate. In: RONCARATI, Cláudia (org.). *Português Brasileiro – contato linguístico, heterogeneidade e história*. Rio de Janeiro: Viveiros de Castro. p. 285-302.

NEVES, Maria Helena de Moura (2000). *Gramática dos usos do português*. São Paulo: UNESP.

NUNES, Jairo (1991). Se apassivador e se indeterminador: o percurso diacrônico no português brasileiro. *Cadernos de Estudos Linguísticos*, Campinas, n.20, p. 33-58, jan/jun.

PAIVA, Maria da Conceição (2003). A variável gênero/sexo. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). *Introdução à sociolinguística*. São Paulo: Contexto. p. 33-42.

PONTE, Vanessa (2006). A indeterminação do sujeito em comunidades rurais afro-brasileiras. *V ELFE – Encontro Nacional sobre língua Falada e Escrita*, Maceió-Al.

RAIMUNDO, Jacques (1933). *O elemento afro-negro na língua portuguesa*. Rio de Janeiro: Renascença.

RODRIGUES, Ayron D. (1993). Línguas Indígenas. 500 anos de descobertas e perdas. *D.E.L.T.A.* v. 9, n. 1, p. 82-104

RODRIGUES, Ayrton D. (1996). As línguas gerais sul americanas. *Papia*. Universidade de Brasília, v. 4, n. 2, p. 6-18.

ROLLEMBERG, Vera *et al.* (1991). Os pronomes pessoais sujeito e a indeterminação do sujeito na norma culta de Salvador. *Estudos Lingüísticos e Literários*. Salvador, n. 11, p. 53-74.

SANKOFF, David (1988). Variable Rules. In: AMMON, Ulrich; DITTMAR, Norbert; MATTHEIR, Klaus (ed). *Sociolinguistics*. New York: Academic Press. p. 119-127.

SANTANA, Neila Maria Oliveira (2006). *A indeterminação do sujeito no português rural do semi-árido baiano*. Dissertação de mestrado. Instituto de Letras, UFBA, Salvador.

SILVA, Jorge Augusto Alves da (2005). *A concordância verbal em terceira pessoa do plural no português popular do Brasil: um panorama social de três comunidades do interior do estado da Bahia*. Tese de doutorado. Instituto de Letras, UFBA, Salvador.

SILVA NETO, Serafim da (1963[1950]). *Introdução ao estudo da língua portuguesa no Brasil*. 2 ed. Rio de Janeiro: INL.

SILVA NETO, Serafim da (1988[1957]). *História da língua portuguesa*. 5 ed. Rio de Janeiro: presence.

SYEA, Anand (1993). Null subject in Mauritian creole and the pro-drop parameter. In: BYRNE, Francis; HOLM, John. *Atlantic meets Pacific*. Amsterdam: John Benjamins. p.91-102.

TARALLO, Fernando (1993). Sobre a alegada origem crioula do português brasileiro: mudanças sintáticas aleatória. In: Roberts, Ian; Kato, Mary (org.). *Português Brasileiro: uma viagem diacrônica*. Campinas: Editora da Unicamp. p. 35-68.

WEINREICH, W.; LABOV, W.; HERZOG, M. (2006 [1975]). *Fundamentos empíricos para uma teoria da mudança lingüística*. São Paulo: Parábola. Tradução de Empirical Foundations for a Theory of the Language Changes.