



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

Rua Barão de Geremoabo, nº147 - CEP: 40170-290 - Campus Universitário Ondina Salvador-BA
Tel.: (71) 263 - 6256 – Site: <http://www.ppgll.ufba.br> - E-mail: pgletba@ufba.br



NEILA MARIA OLIVEIRA SANTANA

***A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS RURAL
DO SEMI-ÁRIDO BAIANO***

**SALVADOR
2006**

NEILA MARIA OLIVEIRA SANTANA

**A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS
RURAL DO SEMI-ÁRIDO BAIANO**

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

Orientadora: Prof^a. Dr^a. Jacyra Andrade Mota

**SALVADOR
2006**

Biblioteca Central Reitor Macêdo Costa - UFBA

S232 Santana, Neila Maria Oliveira.

A indeterminação do sujeito no português rural do semi-árido baiano / por Neila Maria Oliveira Santana. - 2006.
152 f.

Orientadora: Profª Drª Jacyra Andrade Mota.

Dissertação (mestrado) - Universidade Federal da Bahia, Instituto de Letras, 2006.

1. Sociolinguística. 2. Língua portuguesa - Sujeito e predicado. 3. Comunidades agrícolas.
4. Negros - Bahia. I. Mota, Jacyra Andrade. II. Universidade Federal da Bahia. Instituto de Letras.
- III. Título.

CDU - 81'27
CDD - 306.44

TERMO DE APROVAÇÃO

NEILA MARIA OLIVEIRA SANTANA

A INDETERMINAÇÃO DO SUJEITO NO PORTUGUÊS RURAL
DO SEMI-ÁRIDO BAIANO

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Letras e
Linguística, Universidade Federal da Bahia, pela seguinte banca examinadora:

Jacyra Andrade Mota – Orientadora _____
Doutora em Letras (Letras Vernáculas), Universidade Federal do Rio de Janeiro (UFRJ)
Universidade Federal da Bahia

Suzana Alice Marcelino Cardoso _____
Doutora em Letras (Letras Vernáculas), Universidade Federal do Rio de Janeiro (UFRJ)
Universidade Federal da Bahia

Cláudia de Souza Cunha _____
Doutora em Letras (Letras Vernáculas), Universidade Federal do Rio de Janeiro (UFRJ)
Universidade Federal do Rio de Janeiro

Salvador, 27 de abril de 2006.

Aos meus pais Pedro e Lúcia, com amor,
dedico esta conquista.

AGRADECIMENTOS

A Deus, pela minha vida cheia de saúde, de paz e de amor. Dou-Lhe graças por ter-me dado força e coragem para enfrentar os desafios e vencê-los como os tenho até hoje vencido.

À professora doutora Jacyra Andrade Mota, que, mesmo com tantos compromissos já firmados, aceitou-me como sua orientanda, ajudando-me com muita boa vontade e competência na realização desta pesquisa.

Aos meus pais, Maria Lúcia Oliveira Santana e Pedro Figueiredo Santana, pelo amor a mim dedicado, pela paciência da minha ausência e por tudo que fizeram para que eu pudesse ter uma oportunidade como esta.

Aos meus irmãos Lúcio, Júnior e Noeli e ao meu sobrinho Davi, minha gratidão por terem contribuído para que eu me tornasse uma pessoa feliz e pelo apoio no decorrer de todo o curso.

À prima e amiga Lília Santanta, que, desde o início desta dissertação, me fez crer que eu era capaz de realizar essa tarefa.

À professora e amiga Aurelina Ariádine, por sugerir a realização do mestrado ainda no curso de graduação e em quem encontrei incentivo e amizade.

À Professora Eliana Pitombo, pela idéia de um trabalho de tal natureza e pela decisiva contribuição nos primeiros passos desta jornada.

Às amigas Ludinalva e Evani e ao amigo Adelino, por terem acreditado em mim e pelo apoio e incentivo constantes.

A amiga Lúcia Perez – que me fez sua “filha” e em quem sempre encontrei apoio, incentivo, confiança e amizade -, pelas palavras de conforto nos momentos de angústias.

A todos os colegas, professores e funcionários do Programa de Pós-Graduação em Letras e Linguística, pela convivência prazerosa e pelo estímulo constante.

À Norma Lúcia Fernandes e Zenaide Carneiro, por tornarem viáveis os *corpora* desta pesquisa.

À Universidade do Estado da Bahia, pela concessão da Bolsa Programa de Apoio a Capacitação Docente.

Às professoras Odete Menon (UFPR), Maria Alice Godoy (CEFET/PR) e Vilma Maria Cavalcante (UFMA), pela boa vontade de enviarem seus trabalhos para que eu pudesse fundamentar o meu.

A todas as pessoas que, compartilhando o meu ideal, me incentivaram a prosseguir nesta aprendizagem, fossem quais fossem os obstáculos.

RESUMO

Nesta dissertação, analisamos, numa perspectiva sociolingüística, algumas possibilidades de indeterminação do sujeito no discurso oral de comunidades rurais afro-descendentes do semi-árido baiano: Piabas, localizada no município de Caém, e Bananal / Barra dos Negros, comunidades gêmeas localizadas no município de Rio de Contas. Partimos do pressuposto de que estão em uso muito mais formas de indeterminação do sujeito do que descreve a tradição gramatical e que existem fatores lingüísticos e extralingüísticos que condicionam o uso dessas formas. Para esse estudo, consideramos que os seguintes recursos podem conter, em maior ou menor grau, a possibilidade de indeterminar o sujeito: *a gente*, *nós*, *você*, *eles*, $\emptyset+V3PP$, $\emptyset+V3PS$, $\emptyset+VINF$ e $\emptyset+V+SE$. Na análise dos dados, submetidos ao pacote de programas estatísticos VARBRUL, levamos em consideração fatores lingüísticos, como tempo e modo verbal, tipo de verbo, tipo de oração, preenchimento do sujeito, grau de indeterminação, forma antecedente e mudança / manutenção do referente, e extralingüísticos, como gênero, faixa etária e localidade, que pudessem influenciar no comportamento lingüístico do falante. Os resultados mostraram que: a) as formas apresentadas pela gramática tradicional não são as mais utilizadas nos *corpora*, as formas *a gente* e $\emptyset+V3PS$ são as mais empregadas; e b) os fatores lingüísticos e extralingüísticos considerados exercem influência na escolha do recurso que o falante se vale para indeterminar o sujeito de seus enunciados.

Palavras-chave: Indeterminação do Sujeito. Português Rural. Sociolingüística.

ABSTRACT

In this dissertation, we analyze, on a sociolinguistic perspective, some possibilities of what is called indetermination of the subject in sentences of oral Brazilian-Portuguese speech of Afro-Brazilian rural communities, located in the semi-arid region of Bahia State: Piabas, in the municipal district of Caém, and Bananal / Barra dos Negros, twin communities located in the municipal district of Rio de Contas. Our first hypothesis was that many more forms of indetermination of the subject than those described by the Portuguese grammatical tradition were in use in those communities, determined by linguistic and extra-linguistic factors. For this study, we considered the following resources as able to have, in higher or lower degree, the possibility to indeterminate the subject: *a gente*, *nós*, *você*, *eles*, $\emptyset+V3PP$, $\emptyset+V3PS$, $\emptyset+VINF$ e $\emptyset+V+SE$. For the data analysis, submitted to the statistical programs VARBRUL, we took in consideration linguistic factors such as verb tense, kind of verb, kind of sentence, fulfillment of the subject, indetermination degree, preceded form and change / maintenance of the referential; and extra-linguistic factors, such as gender, age level and location, that could influence on the speakers' linguistic behavior. The results showed that, first, the forms described by the traditional Portuguese grammar are not the most used in the corpora and that the forms *a gente* and $\emptyset+V3PS$ are the most employed. Second, that the linguistic and extra-linguistic factors really have influence on the choices speakers make to indeterminate the subject of their sentences.

Key-words: Indetermination of the subject. Rural Portuguese. Sociolinguistic.

LISTA DE TABELAS

Tabela 1 –	Frequência dos recursos de indeterminação do sujeito nos <i>corpora</i>	94
Tabela 2 –	Distribuição dos dados de acordo com as comunidades	96
Tabela 3 –	Influência da forma antecedente no uso de <i>a gente</i>	98
Tabela 4 –	Efeito da forma antecedente e da mudança/ manutenção do referente no uso de <i>a gente</i>	99
Tabela 5 –	Influência do grau de indeterminação no uso de <i>a gente</i>	101
Tabela 6 –	Influência do tempo e modo verbal no uso de <i>a gente</i>	102
Tabela 7 –	Influência da localidade no uso de <i>a gente</i>	103
Tabela 8 –	Efeito da faixa etária e da localidade no uso de <i>a gente</i>	105
Tabela 9 –	Influência do gênero no uso de <i>a gente</i>	106
Tabela 10 –	Efeito do gênero, da faixa etária e da localidade no uso de <i>a gente</i>	107
Tabela 11 –	Influência da forma antecedente no uso de <i>nós</i>	108
Tabela 12 –	Efeito da forma antecedente e da mudança/ manutenção do referente no uso de <i>nós</i>	109
Tabela 13 –	Influência do grau de indeterminação no uso de <i>nós</i>	110
Tabela 14 –	Influência do preenchimento do sujeito no uso de <i>nós</i>	110
Tabela 15 –	Influência da localidade no uso de <i>nós</i>	113
Tabela 16 –	Influência do gênero e da localidade no uso de <i>nós</i>	113
Tabela 17 –	Efeito da faixa etária e da localidade no uso de <i>nós</i>	113
Tabela 18 –	Influência do tempo e modo verbal no uso de <i>nós</i>	114
Tabela 19 –	Influência da variável tipo de verbo no uso de <i>nós</i>	115
Tabela 20 –	Influência da forma antecedente no uso de <i>você</i>	116
Tabela 21 –	Influência do grau de indeterminação no uso de <i>você</i>	116
Tabela 22 –	Efeito da faixa etária e da localidade no uso de <i>você</i>	118
Tabela 23 –	Influência do tempo e modo verbal no uso de <i>você</i>	119
Tabela 24 –	Influência do gênero no uso de <i>você</i>	120
Tabela 25 –	Efeito do gênero e da localidade no uso de <i>você</i>	121
Tabela 26 –	Influência da forma antecedente no uso de <i>eles</i>	122
Tabela 27 –	Influência do grau de indeterminação no uso de <i>eles</i>	123
Tabela 28 –	Influência do tempo e modo verbal no uso de <i>eles</i>	124
Tabela 29 –	Efeito da faixa etária e da localidade no uso de <i>eles</i>	125
Tabela 30 –	Influência da forma antecedente no uso do \emptyset +V3PS	126

Tabela 31 –	Influência do tempo e modo verbal no uso do $\emptyset+V3PS$	127
Tabela 32 –	Influência do grau de indeterminação no uso do $\emptyset+V3PS$	128
Tabela 33 –	Influência do gênero no uso do $\emptyset+V3PS$	128
Tabela 34 –	Efeito do gênero e da localidade no uso do $\emptyset+V3PS$	129
Tabela 35 –	Influência do tempo e modo verbal no uso do $\emptyset+V3PP$	131
Tabela 36 –	Influência da forma antecedente no uso do $\emptyset+V3PP$	132
Tabela 37 –	Influência do grau de indeterminação no uso do $\emptyset+V3PP$	133
Tabela 38 –	Influência da localidade no uso do $\emptyset+V3PP$	133
Tabela 39 –	Influência do gênero e da localidade no uso do $\emptyset+V3PP$	134
Tabela 40 –	Efeito da faixa etária e da localidade no uso do $\emptyset+V3PP$	134
Tabela 41 –	Influência da forma antecedente no uso do $\emptyset+VIN F$	135
Tabela 42 –	Influência do tipo de oração no uso do $\emptyset+VIN F$	136
Tabela 43 –	Influência do grau de indeterminação no uso do $\emptyset+VIN F$	137
Tabela 44 –	Influência do gênero no uso do $\emptyset+VIN F$	137
Tabela 45 –	Efeito da faixa etária e da localidade no uso do $\emptyset+VIN F$	138

LISTA DE QUADROS E GRÁFICOS

Gráfico 1 –	Distribuição dos recursos de indeterminação do sujeito nos <i>corpora</i>	94
Gráfico 2 –	Influência da faixa etária no uso de <i>a gente</i>	104
Gráfico 3 –	Efeito do gênero e da faixa etária no uso de <i>a gente</i>	106
Gráfico 4 –	Influência da faixa etária no uso de <i>nós</i>	111
Gráfico 5 –	Efeito do gênero e da faixa etária no uso de <i>nós</i>	112
Gráfico 6 –	Influência da faixa etária no uso de <i>você</i>	118
Gráfico 7 –	Efeito do gênero e da faixa etária no uso de <i>você</i>	120
Gráfico 8 –	Efeito do gênero e da faixa etária no uso de <i>eles</i>	124
Gráfico 9 –	Influência da faixa etária no uso do $\emptyset+V3PS$	130
Gráfico 10 –	Efeito do gênero e da faixa etária no uso de $\emptyset+V3PS$	130
Gráfico 11 –	Efeito do gênero e da faixa etária na escolha da forma $\emptyset+VINF$	138
Quadro 1 –	Posição dos grupos de fatores selecionados para o uso de cada variante da indeterminação	139

LISTA DE ABREVIATURAS E SIGLAS

Apl	Aplicação
B	Bananal / Barra dos Negros
Cf	Conferir
DID	Diálogo entre informante e documentador
D2	Diálogo entre dois informantes
Doc	Documentador
EF	Elocução formal
FN	Forma nominal
GTs	Gramáticas Tradicionais
Inf	Informante
NELPRU	Núcleo de Estudos em Língua Portuguesa Rural
NURC	Norma Urbana Culta
Ø + V3P + SE	Verbo na terceira pessoa mais a partícula ‘se’ sem sujeito lexical
Ø + V3PP	Verbo na terceira pessoa do plural sem sujeito lexical
Ø + V3PS	Verbo na terceira pessoa do singular sem sujeito lexical
Ø + VINF	Verbo no infinitivo sem sujeito lexical
P	Piabas
P. R.	Peso Relativo
PEUL	Programa de Estudos sobre o Uso da Língua
PORCUFORT	Português Oral Culto de Fortaleza
VARBRUL	<i>Variables Rules</i> (Regras variáveis)
VARSUL	Variação Lingüística Urbana na Região Sul
VPASSINT	Voz passiva sintética
VPSA	Voz passiva sem agente

SUMÁRIO

LISTA DE TABELAS		08
LISTA DE QUADROS E GRÁFICOS		10
LISTA DE ABREVIATURAS E SIGLAS		11
1	INTRODUÇÃO	16
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	A SOCIOLINGÜÍSTICA	19
2.2	A DIALETOLOGIA	26
3	A INDETERMINAÇÃO	35
3.1	ABORDAGEM DA GRAMÁTICA TRADICIONAL	35
3.1.1	Aspectos observados na abordagem gramatical	38
3.1.1.1	Os critérios de análise	38
3.1.1.2	Delimitação dos recursos de indeterminação	40
3.1.1.3	O contexto	41
3.1.1.4	Os conceitos de indefinição e de impessoalização	42
3.2	TRABALHOS SOBRE A INDETERMINAÇÃO EM PORTUGUÊS	44
3.2.1	A indeterminação na língua oral	44
3.2.2	A indeterminação na língua escrita	52
3.3	A INDETERMINAÇÃO SOB A NOSSA PERSPECTIVA	53
3.3.1	Determinação e indeterminação	54
3.3.2	Noção de pessoa	55
3.3.3	Graus de indeterminação	56
3.3.4	Equivalência e alternância	59
3.3.5	Conclusão	60

4	METODOLOGIA	61
4.1	OS <i>CORPORA</i>	61
4.1.1	O projeto “A Língua Portuguesa no Semi-Árido Baiano”	61
4.1.2	As comunidades	62
4.1.3	Os informantes	65
4.1.4	As entrevistas	65
4.1.5	A transcrição	66
4.2	SUPORTE QUANTITATIVO	66
4.3	LEVANTAMENTO DAS OCORRÊNCIAS	68
4.3.1	Critérios de seleção das ocorrências	69
4.4	VARIÁVEIS CONSIDERADAS	70
4.4.1	Variável dependente	70
4.4.1.1	Nós	71
4.4.1.2	A gente	72
4.4.1.3	Você	74
4.4.1.4	Eles	75
4.4.1.5	Ø+V3PS	77
4.4.1.6	Ø+V3PP	78
4.4.1.7	Ø+VINF	79
4.4.1.8	Ø+ V+SE	79
4.4.2	Variáveis independentes	80
4.4.2.1	Variáveis lingüísticas	80
4.4.2.1.1	Forma antecedente	80
4.4.2.1.2	Mudança/ manutenção do referente	84
4.4.2.1.3	Tempo e modo verbal	84
4.4.2.1.4	Tipo de oração	85
4.4.2.1.5	Tipo de verbo	87
4.4.2.1.6	Preenchimento do sujeito	88
4.4.2.1.7	Grau de indeterminação	88
4.4.2.2	Variáveis extralingüísticas	89
4.4.2.2.1	Gênero	89
4.4.2.2.2	Faixa etária	91
4.4.2.2.3	Localidade	91

5	ANÁLISE DOS DADOS	94
5.1	VARIÁVEIS SIGNIFICATIVAS SELECIONADAS PELAS FORMAS DE INDETERMINAÇÃO DO SUJEITO	97
5.1.1	A gente	97
5.1.1.1	Forma antecedente	98
5.1.1.2	Grau de indeterminação	101
5.1.1.3	Tempo e modo verbal	102
5.1.1.4	Localidade	103
5.1.1.5	Faixa etária	104
5.1.1.6	Gênero	105
5.1.2	Nós	107
5.1.2.1	Forma antecedente	108
5.1.2.2	Grau de indeterminação	109
5.1.2.3	Preenchimento do sujeito	110
5.1.2.4	Gênero e faixa etária	111
5.1.2.5	Localidade	112
5.1.2.6	Tempo e modo verbal	114
5.1.2.7	Tipo de verbo	115
5.1.3	Você	115
5.1.3.1	Forma antecedente	116
5.1.3.2	Grau de indeterminação	116
5.1.3.3	Faixa etária	117
5.1.3.4	Tempo e modo verbal	119
5.1.3.5	Gênero	120
5.1.4	Eles	121
5.1.4.1	Forma antecedente	122
5.1.4.2	Grau de indeterminação	122
5.1.4.3	Tempo e modo verbal	123
5.1.4.4	Gênero e faixa etária	124
5.1.5	Ø+V3PS	125
5.1.5.1	Forma antecedente	126
5.1.5.2	Tempo e modo verbal	127
5.1.5.3	Grau de indeterminação	128

5.1.5.4	Gênero	128
5.1.5.5	Faixa etária	129
5.1.6	Ø+V3PP	131
5.1.6.1	Tempo e modo verbal	131
5.1.6.2	Forma antecedente	132
5.1.6.3	Grau de indeterminação	132
5.1.6.4	Localidade	133
5.1.7	Ø+VINFIN	135
5.1.7.1	Forma antecedente	135
5.1.7.2	Tipo de oração	136
5.1.7.3	Grau de indeterminação	136
5.1.7.4	Gênero	137
5.2	ANÁLISE COMPARATIVA	138
6	CONSIDERAÇÕES FINAIS	142
	REFERÊNCIAS	146

1 INTRODUÇÃO

As Gramáticas Tradicionais (GTs) não definem de forma satisfatória a indeterminação do sujeito em português. De modo geral, o sujeito indeterminado é apresentado por elas através de duas formas: a) por um sujeito não preenchido lexicalmente acompanhado de verbo na terceira pessoa do plural, como em (01), e b) pelo pronome “se” junto a verbos na terceira pessoa do singular, como em (02), (03) e (04)¹.

(01) **Contaram-me**, quando eu era pequenina, a história duns naufragos, como nós.

(02) **Desconfiou-se** das propostas.

(03) Ainda **se vivia** num mundo de certezas.

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

Comparando-se as prescrições das GTs e o uso do sujeito indeterminado em diferentes modalidades da língua portuguesa, tanto oral quanto escrita, tanto popular quanto culta, podemos constatar que há uma grande diferença entre o que estas gramáticas prescrevem e o que realmente ocorre no português brasileiro, particularmente o falado.

Os estudos sobre a indeterminação do sujeito no português falado por indivíduos de diferentes regiões do Brasil (cf. MILANEZ, 1982; ALMEIDA, 1992; CUNHA, 1993; MENON, 1994; SETTI, 1997, CAVALCANTE, 1998; GODOY, 1999) têm revelado um quadro de estratégias de indeterminação bem maior do que o apresentado pela GT, como o uso das formas pronominais *eu, tu, você, a gente, nós e eles*, enquanto é cada vez menos freqüente o uso do verbo na terceira pessoa do plural e do verbo na terceira pessoa do singular mais a partícula *se*.

Diante disto, este trabalho tem os seguintes objetivos: a) analisar as possibilidades de indeterminação do sujeito usadas no dialeto de comunidades rurais afro-descendentes do semi-árido baiano: Piabas, localizada no município de Caém, e Bananal / Barra dos Negros, comunidades gêmeas localizadas no município de Rio de Contas; e b) verificar quais são os fatores lingüísticos e extralingüísticos que condicionam o uso de diferentes formas de indeterminação do sujeito no dialeto pesquisado.

¹ Os exemplos foram retirados de Cunha e Cintra (2001, p. 128-9) e de Ferreira (1992, p. 204).

Para observação do fenômeno em estudo, analisaremos as formas pronominais *a gente*, *nós*, *você* e *eles* e as formas sem sujeito lexical $\emptyset+V3PP$, $\emptyset+V3PS$, $\emptyset+VINF$ e $\emptyset+V+SE$ a partir de 24 entrevistas, 12 de cada uma das comunidades acima referidas, que compõem os *corpora* do Projeto “A Língua Portuguesa no Semi-Árido Baiano”.

Nossa hipótese é que estão sendo usados no português, no caso específico desta pesquisa, em comunidades rurais afro-descendentes, diferentes recursos de indeterminação do sujeito que não são apresentados na GT e que tanto os fatores lingüísticos quanto os extralingüísticos influenciam na escolha do recurso utilizado pelo falante para indeterminar o sujeito dos seus enunciados.

Na análise dos dados, faremos uma abordagem variacionista, nos moldes desenvolvidos por Labov (1983)², seguindo os passos da pesquisa sociolingüística (cf. TARALLO, 1990; MOLLICA; BRAGA, 2003). Os dados serão submetidos ao pacote de programas estatísticos VARBRUL para a partir dos resultados verificarmos quais são os condicionamentos lingüísticos e extralingüísticos que podem inibir ou favorecer o uso dos recursos de indeterminação do sujeito analisados neste trabalho.

Organizamos a presente dissertação da seguinte forma:

Após a definição das finalidades deste estudo (nesta introdução), fizemos uma abordagem das teorias que o fundamentam: a Sociolingüística e a Dialectologia, no capítulo denominado *Fundamentação Teórica*.

Em seguida, no capítulo intitulado *A Indeterminação*, apresentamos o fenômeno analisado, a indeterminação do sujeito, descrevemos a abordagem feita pela GT, analisando alguns aspectos propostos por ela e fizemos uma revisão bibliográfica dos trabalhos realizados sobre o tema, tanto da língua oral quanto da língua escrita, centrando-nos no conceito de indeterminação, nos recursos apresentados, nos graus de indeterminação e na análise dos resultados para, a partir dos aspectos expostos pelos trabalhos, chegarmos ao nosso ponto de vista sobre o fenômeno em estudo.

Na *Metodologia*, capítulo seguinte, esclarecemos os passos seguidos no empreendimento desta pesquisa. Descrevemos os *corpora* analisados, o suporte quantitativo, os critérios utilizados na seleção das ocorrências. Nesse capítulo, incluímos ainda a descrição das variáveis dependentes e independentes a que os dados foram submetidos.

² Tradução de *Patterns Sociolinguistics* publicado em 1972 e traduzido em 1983.

No último capítulo, procedemos à análise dos dados. Neste, apresentamos os resultados numéricos obtidos com o auxílio do programa VARBRUL. Os resultados estão organizados a partir da ordem de seleção das variáveis consideradas estatisticamente relevantes para cada uma das formas de indeterminação, estando dispostos em tabelas e gráficos.

Finalmente, nas considerações finais, reunimos as informações mais importantes no estudo do fenômeno variável aqui focalizado.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresentaremos as teorias que fundamentam esta pesquisa. Inicialmente, faremos uma exposição dos princípios básicos da Sociolinguística, que têm como principal representante William Labov, e em seguida faremos uma breve abordagem da Dialetoleologia em suas vertentes tradicional e moderna.

2.1 A SOCIOLINGÜÍSTICA

A Sociolinguística é uma das áreas da Linguística Moderna que estuda a língua em situações reais de uso dentro de uma comunidade de fala, voltando a sua atenção para a investigação dos seus aspectos lingüísticos e sociais. O termo Sociolinguística fixou-se em 1964 num congresso realizado na Universidade da Califórnia em Los Angeles (UCLA), em que se discutiram as relações entre linguagem e sociedade. Desse congresso fizeram parte especialistas de diversas áreas das Ciências Sociais, dentre eles, podemos citar: John Gumperz, Einar Haugen, William Labov, Dell Hymes, John Fisher, José Pedro Rona, conforme apresenta Alkmim (2003a).

Os princípios básicos da Sociolinguística foram publicados primeiramente em 1966, por William Bright, no texto introdutório da obra *Sociolinguistics* que reúne os trabalhos apresentados no congresso acima citado. Nesse texto, Bright propõe que a Sociolinguística deve preocupar-se em sistematizar a variação lingüística observada numa dada comunidade, de acordo com as diferenças sociais desta mesma sociedade, além disso, ele aponta que a identidade social do falante, a identidade social do ouvinte, o contexto da interação verbal e o comportamento social que os falantes fazem do comportamento lingüístico são fatores sociais que interferem na variação lingüística.

O estabelecimento da Sociolinguística é precedido pela atuação de vários pesquisadores que buscavam articular a linguagem com aspectos de ordem social e cultural. Dentre estes pesquisadores está Labov que em 1963, ao realizar o seu trabalho sobre o uso dos ditongos pelos falantes da ilha de Martha's Vineyard, localizada em Massachussetts (EUA), destaca o papel decisivo dos fatores sociais na explicação da variação lingüística. Nesse trabalho, o autor relaciona fatores como sexo, ocupação, origem étnica e atitude ao comportamento lingüístico manifesto dos vineyardenses.

As técnicas desenvolvidas em Martha's Vineyard foram mais bem elaboradas e aplicadas por Labov em outra pesquisa sobre a estratificação social do inglês de Nova York,

realizada em 1964. Ao finalizar esta pesquisa, Labov fixa um modelo de descrição e interpretação do fenômeno lingüístico capaz de dar conta da influência de fatores extralingüísticos no contexto social de comunidades urbanas. Esse modelo ficou conhecido como *Sociolingüística Variacionista* ou *Teoria da Variação*.

Na obra intitulada *Modelos Sociolingüísticos*, Labov (1983) deixa claro que, ao publicar seus primeiros estudos sobre Martha's Vineyard e a cidade de Nova York, na década de 60, a sua intenção era introduzir a contraparte social inerente às línguas, que até então ficara de lado nos estudos realizados. A sua pretensão era observar “el language hablado tal como lo utilizan en la vida cotidiana los miembros de una sociedad organizada; ese vehículo de comunicación en que discuten con sus esposas, juegan con sus amigos y engañan sus enemigos”³ (p. 23).

É consolidada, assim, a Sociolingüística com seu objeto de estudo definido – a variação lingüística – e com a tarefa de “descrever a covariação sistemática entre a estrutura lingüística e a estrutura social” (ALKMIM, 2003b, p. 595).

A variação é percebida como requisito ou condição do próprio sistema lingüístico. Ela é essencial à própria natureza da linguagem humana, pois a língua não é usada de modo homogêneo por todos os falantes. Ao trabalhar com o conceito de variação lingüística, pretendemos demonstrar que a língua portuguesa, assim como todas as línguas do mundo, não se apresenta de maneira uniforme em todo o território brasileiro. O processo de variação ocorre em todos os níveis de funcionamento da língua, seja fonológico, morfológico, sintático, semântico, sendo mais perceptível na pronúncia e no vocabulário.

As variedades observadas na língua, nos estudos sobre variação lingüística, segundo Alkmim (2003a, p. 34), “são relacionáveis a fatores diversos: dentro de uma mesma comunidade de fala, pessoas de origem geográfica, de idade, de sexo diferentes falam distintamente”. Ainda de acordo com Alkmim, as variedades lingüísticas podem ser descritas a partir de dois parâmetros básicos: a **variação geográfica** (ou diatópica) e a **variação social** (ou diastrática).

A **variação geográfica** ou **diatópica** está relacionada às diferenças lingüísticas distribuídas no espaço físico observáveis entre falantes de origens geográficas distintas em que se falam a mesma língua. Essa variação normalmente acontece:

- a) pelas influências que cada região sofreu durante sua formação;

³ A linguagem falada tal como utilizada na vida cotidiana pelos membros de uma sociedade organizada; esse veículo de comunicação em que discutem com suas esposas, jogam com seus amigos e enganam seus inimigos [tradução nossa].

- b) porque os falantes de uma dada região constituem uma comunidade lingüística geograficamente limitada em função de estarem polarizados em termos políticos e/ou econômicos e/ou culturais, sendo assim, desenvolvem um comportamento lingüístico comum que os identifica e distingue.

A **variação social** ou **diastrática** representa um conjunto de fatores que tem a ver com a identidade dos falantes e também com a organização sociocultural da comunidade de fala. Desta forma, podemos apontar alguns fatores relacionados a este tipo de variação:

- a) *classe social* – a classe social a que pertence o indivíduo exerce fortes influências em seu modo de falar. “Há uma tendência para maior semelhança entre os atos verbais dos membros de um mesmo setor sócio-cultural da comunidade, geralmente com relações bastante estreitas de interesses comuns” (CAMACHO, 1988 *apud* TRAVAGLIA, 2002, p. 45). Neste tipo de variação atuam também fatores como nível de escolaridade que está quase sempre inter-relacionado com a classe econômica do falante.
- b) *idade* – a idade representa as variações decorrentes da diferença do modo de usar a língua por pessoas de idades diferentes, normalmente em faixas etárias diversas, tais como, crianças, jovens, adultos e velhos. Durante a vida, a pessoa passa de um grupo para outro, adotando as formas de um grupo e abandonando as do outro.
- c) *sexo* – os estudos mostram que homens e mulheres não falam da mesma maneira. Segundo Monteiro (2000, p. 71),

além das diferenças no ritmo e tom de voz, há preferências por certas estruturas sintáticas, pelo emprego de determinados vocábulos ou fórmulas de cortesia, bem como pela omissão de outros em função das conotações que possam apresentar. Há inclusive certas crenças populares de que as mulheres falam muito mais do que os homens ou que falam bem mais rápido.

Chambers e Trudgill (1994) relatam que as diferenças lingüísticas relacionadas ao sexo ocorrem desde a fala das crianças.

- d) *situação ou contexto social* – representa a variação da fala do indivíduo a depender do contexto em que ele se encontra. Labov (1983) afirma que não existem falantes de estilo único. Há pessoas que apresentam um campo de alternâncias estilísticas mais amplo do que outras, mas “qualquer pessoa muda sua fala, de acordo com o(s) seu(s) interlocutor(es), segundo o lugar em que se encontra e até mesmo segundo o tema da conversa” (ALKMIM, 2003a, p. 36).

Qualquer um de nós é capaz de perceber que a língua está sujeita não só a variações, mas também a mudanças. É uma tendência comprovada empiricamente dizer que as línguas mudam no eixo do tempo. Porém, o complexo jogo que se estabelece entre mudança e permanência não é uma questão que se possa considerar tão simples. Com a publicação, em 1968, do artigo *Empirical Foundations for a theory of language change* escrito por Labov em co-autoria com Weinreich e Herzog apareceu a primeira proposta concreta para tratar a questão da variação e da mudança na língua a partir de fatos lingüísticos observáveis. Esses autores, ao apresentar e discutir essa proposta, levantam algumas questões que uma teoria de base empírica deveria dar conta. Para isso, a base dos questionamentos dos estudos sobre a mudança lingüística encontra-se organizada sob a forma de cinco problemas:

- 1) O *constraint problem* – problema das restrições – remete à questão da definição das condições que favorecem ou restringem a mudança e, por conseguinte, ao conjunto das mudanças possíveis.
- 2) O *embedding problem* – problema de encaixamento – traz para discussão a relação entre mudança e contexto, abordando a relação existente entre fatores internos e externos.
- 3) O *transition problem* – problema de transição – está relacionado à necessidade de definir e analisar o percurso através do qual cada mudança se realiza.
- 4) O *evaluation problem* – problema de avaliação – está relacionado à reação dos falantes diante dos valores da variável observada.
- 5) O *actuation problem* – problema de implementação – esclarece acerca dos fatores responsáveis pela implementação da mudança, ou seja, de como a mudança chegaria a sua consecução.

Para os referidos autores, no entanto, uma teoria da mudança deve lidar com o modo como uma estrutura lingüística de uma comunidade é transformada no curso do tempo, de forma que tanto a língua quanto a comunidade permaneçam as mesmas.

Discorrendo sobre a variação, Monteiro (2000, p. 16) afirma que “os propósitos de descrever a heterogeneidade lingüística e de encontrar um modelo capaz de dar conta da influência dos fatores sociais que atuam na língua somente passaram a ter êxito com os trabalhos de Labov”. Aparentemente caótica e aleatória, a face heterogênea imanente da língua é regular, sistemática e previsível. A sistematização da variação implica a delimitação dos fenômenos variáveis a serem analisados (*variável dependente*) e dos fatores que os regulam (*variáveis independentes*). Assim, a principal tarefa da Sociolingüística é

correlacionar a variável dependente com variáveis independentes, como contexto lingüístico, estilo ou categorias sociais.

Para Mollica (2003, p. 10-1), “a variação lingüística constitui um fenômeno universal e pressupõe a existência de formas lingüísticas alternativas denominadas variantes”. Assim, *variantes* são duas ou mais formas distintas que permitem dizer a mesma coisa num mesmo contexto, configurando um fenômeno variável chamado de *variável dependente*.

Segundo Labov (1983), os falantes não aceitam facilmente o fato de que duas expressões distintas signifiquem exatamente a mesma coisa, havendo, pois, uma forte tendência a conferir-lhes significados diferentes. Se um determinado grupo utiliza uma variante particular, normalmente as conotações sociais atribuídas a esse grupo transferem-se à variável lingüística. Não há problema quando a variação encontra-se no nível fonético, neste nível, as variantes são idênticas quanto à referência ou valor de verdade, mas numa variação morfossintática, por exemplo, é mais difícil sustentar quando duas ou mais estruturas expressam um único significado. “É que a própria hipótese de uma variação sintática por natureza é problemática. As variantes são seqüências que têm o mesmo sentido e comportam os mesmos itens lexicais, mas seus processos combinatórios e hierarquias de categorias são diferentes” (MONTEIRO, 2000, p. 60).

Essa questão do significado tem suscitado algumas críticas relacionadas à aplicação da metodologia variacionista ao estudo de aspectos morfossintáticos (Cf. LAVANDERA, 1984). O argumento utilizado, geralmente, é o de que não se pode prever os contextos de ocorrência das formas alternantes de fenômenos que não sejam fonológicos, já que não há sinônimos perfeitos na língua. Em vista disso, Lavandera (1984) sugere que se enfraqueça a condição de igualdade de significado de todas as formas alternantes no estudo da variação que não seja fonológica, substituindo-a por uma condição igualmente funcional. Para Monteiro (2000, p. 62), no estudo da variação morfossintática “há estruturas que só aparentemente podem ser consideradas variantes (...). O investigador, antes de decidir analisar um fenômeno de variação, deve então ter o cuidado de não incorporar numa mesma variável estruturas de natureza distinta”.

Tendo em vista a questão de que nem toda variável lingüística possa ser objeto de investigação, Labov (1983, p. 36) estabelece algumas condições que devem ser preenchidas antes do pesquisador iniciar o seu trabalho:

En primer lugar, nos interesa un elemento que sea frecuente, que ocurra tan a menudo en una conversación espontánea que su comportamiento pueda ser establecido a partir de contextos no estructurados y entrevistas breves. En segundo lugar, tendría que ser estructural: cuanto más integrado esté el elemento en un sistema mayor de unidades en funcionamiento, mayor será el interés lingüístico intrínseco para nuestro estudio. En tercer lugar, la distribución del rasgo debería estar altamente estratificada: es decir, que nuestras exploraciones preliminares sugerirían una distribución asimétrica mejor que una amplia serie de niveles de edad o cualquier otro estrato ordenado de la sociedad⁴.

No caso específico da variação no uso das formas de indeterminação do sujeito, objeto da presente pesquisa, verificamos a possibilidade de utilização do modelo variacionista. Os diferentes recursos utilizados pelo falante para indeterminar o sujeito de seus enunciados não constituem necessariamente variantes perfeitas no sentido laboviano e tampouco apresentam o mesmo grau de indeterminação⁵. Porém, mesmo que as formas de indeterminação do sujeito possam apresentar nuances semânticas distintas, ocorrem contextos em que é possível o intercâmbio de duas formas diferentes, sem que haja qualquer mudança de significado, como mostram os exemplos (05) e (06) a seguir:

(05) Inf.: E a gente quando chega o dia de votar, a **gente** vota. Agora, ninguém sabe pra quem é que **nós** vota. É isso mermo. (P, 3, 72)⁶

(06) Inf.: Quando **você** entra numa rua aqui, na mesma rua que **a gente** entra, ele pensa que **a gente** vai sair nessa mesma rua, já Ø sai lá adiante. (B, 10, 166)

Há ainda contextos diferenciados em que formas diferentes também têm o mesmo valor de verdade. Podemos comprovar essa afirmação a partir dos exemplos (07) e (08):

(07) Doc.: A senhora acha que o mundo vai se acabar?

Inf.: Não, o mundo não se acaba, é nós, uns matano, os'outro chega, atira. **Nós** tamo aí deitado, a morte pica lo purrete e leva. (P, 12, 173)

⁴ Em primeiro lugar, nos interessa um elemento que seja freqüente, que ocorra muitas vezes em uma conversa espontânea que seu comportamento possa ser estabelecido a partir de contextos não estruturados e entrevistas breves. Em segundo lugar, teria que ser estrutural: quanto mais integrado estiver o elemento em um sistema maior de unidades em funcionamento, maior será o interesse lingüístico intrínseco para nosso estudo. Em terceiro lugar, a distribuição do deveria estar altamente estratificada: isto é, que nossas explorações preliminares sugeririam uma distribuição assimétrica melhor que uma ampla série de níveis de idade ou qualquer outro estrato ordenado da sociedade [tradução nossa].

⁵ Cf. item 3.3.3: graus de indeterminação.

⁶ As referências aos *corpora* serão feitas como se segue: identificação da comunidade, número do informante e página do inquérito da qual o contexto foi retirado.

(08) Inf.: Hoje em dia **Ø não pode** ficar junto não, tem que a coisa... **Ø tem** que lutar pro mundo. (B, 33, 245)

Nesses exemplos, tanto a forma *nós* quanto a forma $\emptyset+V3PS$ referem-se a ‘qualquer pessoa’.

Definida a *variável dependente* e as condições para que ela ocorra, é oportuno dizer que o emprego de suas variantes não é aleatório, mas influenciado por grupos de fatores de natureza social e/ou lingüística chamados de *variáveis independentes*. Assim, as *variáveis independentes* podem ser de natureza interna ou externa à língua e podem exercer pressão sobre os usos, inibindo ou favorecendo o emprego de formas variantes.

No conjunto das variáveis internas, encontramos os fatores fonológicos, morfológicos, sintáticos, semânticos, discursivos e lexicais que dizem respeito a características da língua em várias dimensões, desde os níveis do significado e do significante, até os seus diversos subsistemas.

No conjunto das variáveis externas à língua reúnem-se os fatores inerentes ao indivíduo, como etnia e gênero, os propriamente sociais, como escolarização, nível de renda, profissão e classe social, e os contextuais, como grau de formalidade e tensão discursiva. Nesse conjunto, registramos ainda os marcadores regionais predominantes em comunidades facilmente identificadas geograficamente.

Vale lembrar, que, como diz Mollica (2003, p. 12), “a complexidade dos condicionamentos da variação não permite a previsão de todos os tipos de agente correlacionados às variantes lingüísticas”, pois cada fenômeno exigirá uma rede de fatores que interage na variação lingüística.

Na análise empreendida, consoante a perspectiva da *Sociolingüística Variacionista*, a correlação entre variáveis dependentes e variáveis independentes é obtida mediante o tratamento quantitativo dos dados, o que é feito com a utilização de um programa estatístico⁷ (o pacote VARBRUL) que trata das regras variáveis.

Resta ainda demonstrar como o modelo variacionista superou o conceito estruturalista de variantes livres. Para os estruturalistas, a variação não seguia qualquer condicionamento e o falante teria a liberdade de escolher qualquer variante que estivesse disponível. Chambers e Trudgill (1994, p. 187) demonstraram que tanto a Sociolingüística como a Dialectologia se opõem a esse princípio, visto que, atualmente, ambas compartilham com vários recortes metodológicos, como os aspectos sociais envolvidos na variação lingüística:

⁷ Falaremos sobre o pacote de programas VARBRUL, suporte quantitativo utilizado neste trabalho, no item 4.2.

La hipótesis de la variación libre sostiene que los elementos variables de habla son apariciones impredecibles de algún rasgo lingüístico que una otra razón, tiene o más realizaciones posibles. Cuando se postula una razón, normalmente implica que el rasgo está en proceso de cambio y que por tanto su representación en la gramática no está asentada. De hecho las variables estudiadas hasta ahora sugieren que no existe tal variación libre, y que los rasgos que varían están condicionados, a veces, por conjunto de factores lingüísticos y sociales⁸.

Se consideramos que a variação é um fenômeno livre, naturalmente não podemos estudá-la como fizeram os estruturalistas. Por essa razão, o modelo variacionista mostra que a variação é estruturada de acordo com as propriedades sistêmicas das línguas e se implementa porque é contextualizada com regularidade.

2.2 A DIALETOLOGIA

Baseada no princípio de que toda língua se caracteriza pela “unidade na diversidade” e pela “diversidade na unidade”, a Dialetoлогия, segundo Rossi (1980, p. 3298), “se propõe inventariar, sistematizar e interpretar as variantes de uma língua, ou de um grupo de línguas definido por qualquer afinidade entre elas, com especial atenção a distributividade – espacial, cronológica, sociocultural etc. – dos traços lingüísticos depreendidos”.

A Dialetoлогия surge no final do século XIX, antes mesmo de os estudos sociolingüísticos se iniciarem, como resposta a uma afirmação teórica dos Neogramáticos⁹. Ao questionar alguns pressupostos tradicionais da prática histórico-comparativa, que fazia estudos entre as línguas clássicas, como o latim, o grego, o sânscrito, o persa, dentre outras, a fim de chegar à língua original indo-européia, esses estudiosos estabeleceram uma orientação metodológica diferenciada e um conjunto de postulados teóricos para a interpretação da mudança lingüística.

Para eles, as mudanças fonéticas estavam subordinadas a uma regularidade absoluta, ou seja, todas as mudanças fonéticas eram governadas por uma regra, afetando sempre a mesma unidade fônica, no mesmo ambiente, em todas as palavras, não admitindo exceções. Se houvesse alguma exceção, seria ou porque o princípio regular efetivo ainda era

⁸ A hipótese da variação livre sustenta que os elementos variáveis da fala são aparições imprescindíveis de algum fenômeno lingüístico que por uma outra razão tem mais realizações possíveis. Quando se postula uma razão, normalmente implica que o fenômeno está em processo de mudança e que, portanto, sua representação na gramática não está estável. De fato, as variáveis estudadas até agora sugerem que não existe tal variação livre, e que os fenômenos que variam estão condicionados, às vezes, por um conjunto de fatores lingüísticos e sociais [tradução nossa].

⁹ Grupo de pesquisadores da Universidade de Leipzig, na Alemanha, conhecidos como *Junggrammatiker* (jovens gramáticos).

desconhecido, ou porque a regularidade da mudança havia sido afetada pelo processo da analogia.

Segundo Faraco (2005, p. 144), para esses estudiosos,

a mudança por analogia era entendida como uma interferência do plano gramatical no plano fônico, o que afetava, em consequência, o caráter absoluto da mudança sonora e criava irregularidades. Ao regularizar gramaticalmente as formas, a analogia romperia a regularidade da mudança fonética.

Com o crescente número de pesquisas comparativas entre as línguas, o número de exceções era ampliado a cada avanço da comparação. Em vista disso, para esses pesquisadores, as irregularidades que ocorriam constantemente nas comparações lingüísticas “se explicariam pelo fato de as línguas investigadas serem resultado daquilo a que hoje se chamaria contatos e interferências (na época, ‘misturas’, ‘contaminações’, ou equivalentes)” (ROSSI, 1980, p. 3298).

Para os estudiosos do século XIX, seria necessário encontrar “as línguas puras”, ou seja, aquelas resultantes de um desenvolvimento histórico imune a “misturas” e “contaminações” e, por isso, as mais aptas a comprovar a regularidade em que se acreditava. Mas, para isso, seria necessário recorrer à língua falada, pois só assim haveria manifestações lingüísticas concretas do povo.

Em consequência disso, os estudos posteriores tiveram como objeto o estudo das variedades lingüísticas usadas por falantes sem escolaridade, principalmente aqueles que pertencessem a comunidades isoladas, que não sofressem influências de outras comunidades.

Esse pensamento contribuiu para o desenvolvimento de um conjunto de métodos de recolha das diferenças dialetais de modo mais sistemático.

Cardoso (2001, p. 27) diz que

o “começo feliz” para a Dialectologia, no seu espectro mais amplo, vem a ter dois marcos que imprimem as primeiras, e principais, diretrizes para trabalho de tal natureza: o levantamento de dados da realidade alemã feito por Wenker e a recolha sistemática para o *Atlas Linguistique de la France* (ALF), obra de Gilliéron e Edmont.

Wenker, a fim de encontrar nos dialetos alemães a confirmação dos princípios dos neogramáticos, elaborou um questionário com uma lista de frases curtas escritas em alemão padrão e enviou, em 1976, para os inspetores regionais de escolas para que estes passassem aos instrutores (professores) para serem traduzidos no dialeto local. Não há no trabalho de Wenker o controle sistemático de variáveis sociais, como o gênero e a faixa etária; quanto à escolaridade, os questionários foram respondidos por pessoas letradas, responsáveis pelo ensino na região. Além disso, a coleta dos dados foi feita por correspondência, tendo profundas implicações para o tratamento de informações fonéticas.

Mesmo assim, o trabalho de Wenker contribuiu para um avanço significativo da Dialectologia, pois com a documentação de fatos de regiões distintas seria possível a interrelação dos dados. A comparabilidade desses dados expostos espacialmente vai-se constituir, segundo Cardoso (2001, p. 27), “numa das prerrogativas dos estudos dialetais”.

Os primeiros resultados da recolha de Wenker foram publicados em 1881, em Strassburg, num trabalho intitulado *Sprachatlas des Deutschen Reichs*, com um conjunto de seis cartas, duas fonéticas e quatro morfológicas, que constituiu o fascículo inicial do *Sprachatlas Von Nord – und Mitteldeutschland, auf Grund Von systematisch mit Hilfe der Volksschuler gesammeltem Material aus circa 30.000 Orten*¹⁰.

Apesar de o trabalho de Wenker ser o marco inicial da geografia lingüística na Alemanha, abrindo caminho para uma ampla pesquisa diatópica, o método de investigação geográfica dos dialetos só é consolidado a partir da publicação do *Atlas Linguistique de la France* (ALF), por Jules Gilliéron, de 1902, quando publica o primeiro volume, a 1910, quando publica o décimo terceiro e último volume, em Paris.

Durante a coleta de dados, Gilliéron aperfeiçoou a metodologia empregada por Wenker baseando-se na documentação *in loco*, o que este não fazia. O responsável pela recolha foi Edmond Edmont, um comerciante da cidade de Saint-Pol que percorreu os campos franceses entre os anos de 1886 e 1900, selecionando informantes e realizando as entrevistas. A técnica de recolha dos informantes não é pré-determinada nem registrada nas cartas do ALF, embora seja depreensível a partir do exame do perfil dos informantes: as idades se estendem dos 15 aos 85 anos, o número de informantes do gênero feminino é menor do que os do gênero masculino e o nível de escolaridade é identificado por dedução das categorias de informantes (CARDOSO, 2001, p. 28).

¹⁰ Atlas lingüístico da Alemanha do norte e do centro, com base em material reunido sistematicamente com a ajuda de professores primários de cerca de 30.000 localidades.

Assim, como dissemos anteriormente, a obra de Gilléron teve o mérito de marcar o início da aplicação do método da Geografia Lingüística, seguindo o rigor científico. A consagração desse método se dá a partir da publicação do *Sprach – und Sachatlas Italiens und der Südschweiz* (*Atlas Ítalo-Suíço - AIS*), por Karl Jaberg e Jakob Jud, em 1928. Outros atlas românicos foram publicados, mas este foi um dos mais importantes, apresentando uma grande inovação em relação ao ALF: explorar a relação entre a palavra e a coisa denominada, avançando no campo das relações etnolingüísticas.

Após a publicação do AIS, o *Linguistic Atlas of New England* (*Atlas da Nova Inglaterra*) foi publicado por Hans Kurathe e seus colaboradores, entre 1939 e 1943. De acordo com Cardoso (2001, p. 30), a grande novidade deste atlas “está no estabelecimento de critérios para escolha dos informantes, envolvendo o aspecto geográfico e o social”. Na elaboração deste atlas, foram considerados três níveis de escolaridade somados a duas faixas etárias, começando, assim, a distribuição sistemática dos fatores sociais, o que não acontecia na metodologia adotada na elaboração dos atlas anteriores.

O controle das variáveis sociais na escolha dos informantes estabelecido na elaboração do *Atlas da Nova Inglaterra* traz novas perspectivas para a Geografia Lingüística, pois este atlas consiste na primeira tentativa de oferecer indicações mais precisas sobre os aspectos sociais da linguagem, abrindo, assim, o caminho para uma nova Dialetologia: a pluridimensional.

Com relação aos estudos dialetais no Brasil, Cardoso (1999, p. 2) afirma que

a primeira manifestação que se pode caracterizar de natureza dialetal sobre o português do Brasil deve-se a Domingos Borges Barros, Visconde de Pedra Branca, que escreveu, em 1926, quando ministro plenipotenciário do Brasil na França e a pedido do geógrafo vêneta Adrien Balbi, um informe sobre *‘les différences que le dialecte brésilien pourrait présenter, comparé à la langue du Portugal’* o qual vem inserido às páginas 172-175 da *Introduction à l'Atlas ethnographique du globe*. Trata-se de uma lista de palavras que apresenta um rol de oito nomes que mudam de significação e outro de cinquenta nomes usados exclusivamente no Brasil.

A produção no campo da Geografia Lingüística ou Geolingüística, no que se refere à realização de atlas lingüísticos, conta hoje com sete publicações em diferentes regiões do Brasil (sul, sudeste e nordeste) e outros em diferentes etapas de andamento.

O primeiro atlas lingüístico produzido no Brasil foi o *Atlas Prévio dos Falares Baianos* (*AFPB*), publicado entre os anos de 1960 e 1963, por Nelson Rossi e as co-autoras Carlota Ferreira e Dinah Maria Isensee. Recobre todo o Estado da Bahia, com uma rede de 50

localidades, que se distribuem pelas diferentes áreas geográficas e culturais. Tem um total de 100 informantes, 57 mulheres e 43 homens, com idade entre 25 e 60 anos. Com relação à escolaridade, todos são analfabetos ou semi-analfabetos. Além dos dados lingüísticos, este atlas traz em suas cartas dados etnográficos.

Em 1977, o primeiro volume de um *Esboço de um Atlas Lingüístico de Minas Gerais (EALMG)* foi publicado, tendo como autores José Ribeiro, Mário Zágari, José Passini e Antônio Gaio. Este atlas foi concebido em quatro volumes dos quais se publicou o primeiro, estando os demais no prelo. A metodologia utilizada valeu-se de inquéritos aplicados diretamente nos pontos selecionados e de inquéritos por correspondência. Quanto à seleção dos informantes, não se restringe ao gênero masculino, adultos, rurícolas e analfabetos, abrindo a possibilidade de análise de outros níveis de variação lingüística.

O *Atlas Lingüístico da Paraíba (ALPB)* foi o terceiro atlas publicado e teve como autoras Maria do Socorro Silva e Aragão e Cleusa Bezerra de Menezes. Está concebido em três volumes dos quais os dois primeiros foram editados em 1984. Em cada localidade foram inquiridos um mínimo de três informantes e um máximo de dez, todos eles com idade compreendida entre 30 e 75 anos.

Embora com dificuldades de financiamento, em 1987 foi publicado o primeiro *Atlas Lingüístico de Sergipe (ALS D)*, executado por um grupo de pesquisadores da Bahia: Carlota Ferreira, Jacyra Mota, Judith Freitas, Nadja Andrade, Suzana Cardoso, Vera Rollemberg e Nelson Rossi. Este atlas segue a mesma recolha de dados e preparação das cartas do AFPB e teve seus originais prontos para impressão desde 1973. Os atlas da Bahia e o de Sergipe seguiram uma mesma orientação metodológica, com algumas modificações em Sergipe. Foram selecionados dois informantes de cada localidade, pertencentes a ambos os gêneros, com escolaridade variando entre analfabetos e alfabetizados e com idade predominantemente entre 35 e 53 anos.

Ainda há, dentre os atlas publicados, o *Atlas Lingüístico do Paraná (ALPr)*, de autoria de Vanderci de Andrade Aguilera, apresentado em 1990 como tese de doutorado e depois publicado em 1994, em dois volumes. Foram pesquisadas 65 localidades distribuídas por todo o Estado do Paraná e ouvidos dois informantes em cada ponto da rede, com idade compreendida entre 30 e 60 anos e com dois graus de escolaridade: primário completo e analfabeto.

O *Atlas Lingüístico-Etnográfico da Região Sul*, coordenado por Walter Koch com a colaboração de Mário Klassmann, José Luiz Mercer, Oswaldo Furlan e Hilda Vieira, foi publicado em 2002. Tem uma característica inovadora por ser o primeiro atlas a não se limitar

ao mapeamento de um estado, ele recobre os três estados da região Sul do país: Rio Grande do Sul, Santa Catarina e Paraná. Sua rede de pontos conta com 294 localidades, sendo 106 no Paraná, 86 em Santa Catarina e 102 no Rio Grande do Sul. Os informantes têm idade entre 28 e 58 anos e pouca escolaridade, sendo dois por localidade nas áreas rurais e três nas áreas urbanas.

O último atlas regional publicado no Brasil foi o *Atlas Lingüístico de Sergipe II (ALS II)*, de autoria de Suzana Alice Marcelino Cardoso, apresentado como tese de doutoramento pela Universidade Federal do Rio de Janeiro em 2002 e publicado em 2005. Este é o segundo atlas de Sergipe e nele é utilizado o *corpus* não explorado no primeiro volume, publicado em 1987. Neste atlas, a autora busca explorar outras variáveis lingüísticas não contempladas no ALS-I, como a diagenérica. A rede de pontos constitui-se de 15 localidades distribuídas por todas as microrregiões homogêneas do estado, contando, cada ponto, com dois informantes, uma mulher e um homem, que foram escolhidos conforme os critérios configurados como básicos para os estudos dialetais – nascidos na cidade objeto de estudo, filhos preferentemente de pais da mesma localidade, não-alfabetizados ou semi-alfabetizados, com afastamento nulo ou por pouco tempo do ponto de residência.

No Brasil existem ainda alguns atlas sendo desenvolvidos que se encontram em diferentes estágios de elaboração:

- O *Atlas Lingüístico do Acre*
- O *Atlas Lingüístico do Ceará*
- O *Atlas Lingüístico do Maranhão*
- O *Atlas Lingüístico do Mato Grosso*
- *Atlas Lingüístico do Mato Grosso do Sul*
- *Atlas Geo-Lingüístico do Pará*
- *Atlas Lingüístico do Rio Grande do Norte*
- O *Atlas Etnolingüístico dos Pescadores do Estado do Rio de Janeiro*
- *Atlas Lingüístico de São Paulo*

A esse conjunto de atlas regionais, situa-se o Projeto *Atlas Lingüístico do Brasil* – Projeto ALiB –, projeto de construção de um atlas nacional iniciado em 1996 durante o Seminário “Caminhos e Perspectivas para a Geolingüística no Brasil”, dirigido por um Comitê Nacional constituído por Suzana Alice Marcelino Cardoso, Jacyra Mota, Maria do Socorro Aragão, Mário Zágari, Vanderci Aguilera, Abdelhak Razky, Aparecida Negri Isquerdo, Cleo Altenhofen e Walter Koch. Esse projeto retoma a idéia de realização de um atlas lingüístico geral do Brasil lançada em 1952 por filólogos e lingüistas brasileiros, como Serafim da Silva

Neto, Celso Cunha e Antenor Nascentes, e no desejo do próprio Governo que, ao definir as finalidades da Comissão de Filologia da Casa de Rui Barbosa, manifestou o seu interesse por um atlas nacional, que até 1996 não havia sido implantado.

Foi implantado, então, o Projeto ALiB em 1996 com os seguintes objetivos gerais:

- Descrever a realidade lingüística do Brasil, no que tange à língua portuguesa, com enfoque na identificação das diferenças diatópicas, diastráticas e diageracionais (fônicas, inclusive prosódicas, morfossintáticas, léxico-semânticas), consideradas na perspectiva da Geolingüística pluridimensional.
- Oferecer aos estudiosos da língua portuguesa (lingüistas, lexicólogos, etimólogos, filólogos e das demais áreas dos estudos lingüísticos), aos pesquisadores de áreas afins (história, antropologia, sociologia) e aos pedagogos (gramáticos, autores de livros-texto para o ensino fundamental e básico, professores) subsídios para o aprimoramento do ensino/ aprendizagem e para uma melhor interpretação do caráter multidialetal do Brasil (CARDOSO; MOTA, 2003, p. 40-1).

Em relação à metodologia adotada na coleta de dados, “embora priorize a dimensão diatópica, como se requer de trabalhos desse tipo, o Projeto ALiB busca a inclusão de novos parâmetros como o diageracional, o diagenérico ou diassexual, o diastrático, o diafásico e o diarreferencial” (MOTA, 2002b, p. 365).

A variação diatópica desse atlas é depreendida a partir de uma rede de 250 pontos que se distribuem em sentido norte-sul do país. Diferentemente do que era feito nos estudos dialetológicos, o ALiB não se restringe a um único tipo de informante, considerando

indivíduos dos dois gêneros, sistematicamente distribuídos – 552 homens e 552 mulheres -, de duas faixas etárias distanciadas entre si pelo menos 20 anos – uma mais nova, de 18 a 30 anos, e uma mais velha, de 50 a 65 anos – e, nas capitais de estado, de distintos graus de escolaridade – quatro informantes alfabetizados, com, no máximo, os quatro primeiros anos do curso básico, e quatro informantes de formação universitária (MOTA, 2002b, p. 368).

A partir dessas características são depreendidas a variação diastrática, a diagenérica e a diageracional. A variação diafásica é depreendida a partir do confronto entre os diversos tipos de registros coletados dos informantes, como: as respostas das indagações do inquiridos; as elocuições mais descontraídas e mais coloquiais; a partir da leitura de uma versão adaptada da “Parábola dos Sete Vimes”; e da descrição de gravuras (cf. MOTA, 2002b, p. 369-70). A variação diarreferencial pode ser depreendida nas respostas das questões de natureza metalingüística constantes no questionário.

Sendo assim, o Projeto AliB atende aos princípios da geografia lingüística e introduz controles de natureza sociolingüística, procurando alcançar uma natureza multidimensional da língua portuguesa no Brasil.

A partir desta breve apresentação dos estudos dialetais, vemos que a Dialetoлогия vem passando por avanços graduais, principalmente no que diz respeito à metodologia empregada. Como dizem Oliveira e Isquerdo (2003, p. 50),

os estudos dialetológicos têm acompanhado as exigências das mudanças metodológicas, visando aos diversos procedimentos das perscrutações científicas. Desse modo, levantamentos realizados na elaboração de atlas lingüísticos apontam para resultados mais precisos, sistemáticos e uniformes.

Essas transformações porque a geolingüística vem passando não são recentes, sendo já discutidas desde 1991 no simpósio *Neue Wege der romaischen Geolingustik*, em Heidelberg e Mainz, na Alemanha. Ao falar das inovações da geolingüística moderna, Radtke e Thun (1996, p. 31) dizem que “la ampliación más frecuente se realiza a partir de la combinación del parámetro diatópico con el diastrático y el diageracional¹¹”, assim como os parâmetros diassexual e diafásico, tão pouco considerados na elaboração dos atlas lingüísticos.

Desta forma, como diz Mota (2002b, p. 1),

em sua feição moderna, a Geolingüística substitui a busca de dialetos arcaizantes, isolados, acantonados em localidades antigas e afastadas dos centros maiores, documentados a indivíduos analfabetos, em geral do sexo masculino, ligados ao meio rural, de onde não se tenham afastado, para procurar deprender e sistematizar a complexa rede de variantes que se encontra em todas as línguas, tanto nas pequenas comunidades quanto nas grandes cidades.

A Dialetoлогия Tradicional consistia no estudo de formas lingüísticas geograficamente diferentes em área predominantemente rural. Em sua vertente mais moderna, está sendo substituída pela Dialetoлогия Pluridimensional, acompanhando as mudanças sociopolíticas que as comunidades lingüísticas têm passado, focalizando situações de contato e a mobilidade demográfica das áreas rurais para os grandes centros urbanos.

A Dialetoлогия atingiu o seu estágio atual devido à revisão feita em sua metodologia, não se limitando apenas ao estudo da variação diatópica (geográfica), nem a um único estrato

¹¹ A ampliação mais freqüente se realiza a partir da combinação do parâmetro diatópico com o diastrático e o diageracional [tradução nossa].

social, nem a grupos fixados em zonas rurais. Ao falar das *Perspectivas para a Dialectologia no Brasil*, Cardoso (1996, p. 110) afirma que

o caráter extensivo da Dialectologia, no sentido de que tem predominantemente uma abrangência espacial, diatópica, à qual se podem e devem agregar incursões sistemáticas de natureza sociolingüística – estráticas, etárias, diafásicas -, garante a sua atualidade e assegura a importância da expansão dos estudos nesse campo, pelo menos até que se tenha descrito todo o território brasileiro.

Verificamos então que da união entre Dialectologia e Sociolingüística podemos fazer um estudo mais profundo da língua. Embora os geolingüistas, nas últimas décadas, ainda priorizem o estudo da variação diatópica ou horizontal, eles têm atentado para a análise e incorporação das variações diastrática e diafásica em seus trabalhos. Ao mesmo tempo, examinam os avanços da Sociolingüística e, cada vez mais, refletem sobre a influência do impacto das mudanças sociais do mundo moderno na pesquisa geolingüística. Essas tendências, visíveis nos últimos atlas publicados ou em andamento no Brasil e no exterior, têm possibilitado a rediscussão e reformulação de vários elementos, como:

- a) critérios para a seleção dos pontos de inquérito;
- b) número e características dos sujeitos;
- c) instrumento para a coleta dos dados;
- d) análise e apresentação dos dados.

Desta forma, essas tendências, em gérmen na reflexão sobre a trajetória da chamada geolingüística tradicional e no aporte de outros domínios da atualidade, têm provocado o aparecimento de novas vertentes nos estudos geolingüísticos contemporâneos.

3 A INDETERMINAÇÃO

Neste capítulo, iremos relatar diferentes abordagens sobre a indeterminação, iniciando pela normativa. Faremos uma descrição de como a tradição gramatical concebe a indeterminação do sujeito, no que diz respeito à definição e aos recursos apresentados. Para isso, analisaremos os seguintes autores: Bechara (1992), Cegalla (1995) Cunha e Cintra (2001), Kury (1989), Luft (1996), Melo (1967, 1980) e Rocha Lima (1998). Este procedimento terá como objetivo apontar alguns problemas relacionados à abordagem da GT sobre a indeterminação do sujeito e, a partir daí, nos posicionarmos em relação ao fenômeno em estudo.

Após essa abordagem, faremos uma revisão das pesquisas realizadas sobre o tema sob diferentes enfoques teóricos, com base tanto na modalidade escrita quanto na falada, a fim de mostrar como se realiza a indeterminação do sujeito por falantes de diferentes regiões brasileiras. Serão, então, examinados os seguintes trabalhos: Milanez (1982), Tupiná (1984), Cunha (1993), Lopes (1993), Menon (1994), Machado (1995), Setti (1997), Cavalcante (1998), Cavalcante (1999) e Godoy (1999).

Por fim, a partir dessas abordagens, tentaremos delimitar o conceito de indeterminação.

3.1 ABORDAGEM DA GRAMÁTICA TRADICIONAL

De modo geral, os autores das GTs conceituam o sujeito indeterminado como aquele que não podemos identificar por desconhecermos quem executa a ação ou pelo desinteresse em conhecê-lo, conforme se depreende nas definições abaixo:

Para Melo (1980, p. 193),

o que torna indeterminado um sujeito é a intenção ou a situação do falante, que não sabe ou não quer individualizar, precisar, apontar o agente, o autor da ação ou da façanha.

Cunha e Cintra (2001, p. 125) entendem que:

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.

Do mesmo modo, Bechara (1992, p. 200) afirma que:

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

No que diz respeito à delimitação dos recursos, os processos pelos quais o sujeito indeterminado se realiza são restritos. Em geral, a tradição gramatical aponta apenas duas maneiras de se indeterminar o sujeito em português, como mostram Cunha e Cintra (2001, p. 125):

Nestes casos em que o sujeito não vem expresso na oração nem pode ser identificado, põe-se o verbo:

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

Outras formas são mencionadas, porém, de modo esporádico, por um ou outro autor. Bechara (1992, p. 200), além de indicar que se indetermina o sujeito das duas maneiras acima citadas, diz existir a possibilidade de pôr “o verbo da oração na terceira pessoa do singular, sem referência a pessoa determinada” (geralmente na expressão ‘diz que...’), apesar de ser empregado mais freqüentemente com o verbo na terceira pessoa do plural, sendo as duas construções tidas pelo autor como equivalentes. Sobre isto, o autor cita os seguintes exemplos:

(09a) “**Diz** que eles vão bem”. (diz = dizem)

(09b) “**Dizem** que eles vão bem”. (p. 200)

Kury (1989), numa observação em nota de rodapé, também afirma a equivalência dessas formas, com relação ao verbo ‘dizer’, como expressão da indeterminação do sujeito. Melo (1967) atribui esse uso às linguagens coloquial e literária antiga. Já Luft (1996) faz uma observação em relação à construção ‘Diz que...’ alegando o seu uso, também na linguagem coloquial e na literária antiga, como uma provável redução haplológica¹² de ‘diz-se que’.

Cegalla (1995, p. 296-7) acrescenta às duas formas clássicas outro recurso de indeterminação:

Deixando-se o verbo no infinitivo impessoal.

Exemplos: Era penoso **carregar** aqueles fardos enormes.

É triste **assistir** a estas cenas repulsivas.

¹² Relativo a haplogogia: simplificação por meio de supressão da primeira de duas sílabas iguais ou semelhantes (Cf. Melhoramentos, 1995).

Luft (1996) omite a forma canônica do verbo na terceira pessoa do singular acompanhada do pronome ‘se’ e acrescenta, assim como Cegalla, o uso do verbo no infinitivo impessoal como mais um recurso para indeterminar o sujeito, conforme podemos ver na citação abaixo:

A indeterminação do sujeito se pode fazer de duas maneiras:

1) emprego do verbo na 3ª pessoa do plural: **dizem** que ele vem; **comentam** que ele não volta mais.

2) emprego do verbo no infinitivo dito “impessoal”: é fácil **protestar**; é preciso **manter** a calma; **reclamar** não adianta; ouvi **chamar** da rua. (p. 25)

Melo (1967, p. 44) acrescenta às duas formas tradicionais a possibilidade de se exprimir o sujeito indeterminado por meio de “sujeito materialmente constituído por pronomes indefinidos que nada esclarecem quanto à identidade do agente (...) e pela expressão ‘a gente’, de valor indefinido”, como mostra nos exemplos:

“**Quem** fêz mal ao guerreiro branco na terra dos tabajaras?...

- **Ninguém** fêz mal a teu hóspede, filha de Araquém.” (ALENCAR, *Iracema*, 1948).

“Ele lhe aparecia agora como um dêsses recantos da mata, próximo a um riacho, num sombrio misterioso e confortante. Passando num meio-dia quente, ao trote penoso do cavalo, **a gente** pára ali, olha a sombra e o verde como se fosse para um cantinho do céu...” (QUEIRÓS, *Três Romances*, 1948).

Melo (1967) e Kury (1989) apresentam a forma *a gente* como uma possibilidade de indeterminar o sujeito, porém especificam que ela é uma ‘expressão indefinida’ ou ‘locução pronominal indefinida’. Kury (1989) aponta, ainda, a possibilidade de indeterminação com verbos na primeira pessoa do plural ou na segunda pessoa do singular. Ao apresentar o exemplo (10),

(10) “Estávamos no início da segunda metade do século passado”. (p. 26),

o autor explica que aparentemente o sujeito desta oração é *nós*, que corresponde a ‘estávamos’; mas, na realidade, quando esta forma foi usada, quem a usou não estava presente no acontecimento narrado e ao usá-la quis referir-se a um número de pessoas impossível de se determinar. Desta forma, o autor diz estar diante de um sujeito indeterminado.

Os autores citados apontam ainda a forma *você* para indicar o sujeito indeterminado, fazendo referência não à segunda pessoa, mas a qualquer pessoa, vagamente. Porém especificam¹³ que esta forma é usada apenas no uso coloquial: “contando uma história, descrevendo uma viagem bonita, o falante diz coisas assim: Aí **você** descortina [isto é, **descortina-se**] uma paisagem deslumbrante!” (MELO, 1967, p. 46).

3.1.1 Aspectos observados na abordagem gramatical

A partir da descrição feita, observamos que a indeterminação não vem sendo definida e conceituada de forma satisfatória pelas GTs. De modo geral, as análises sobre a indeterminação do sujeito são diversificadas entre os gramáticos, sendo tratada de forma simplista e superficial. Em relação ao português falado, apenas um ou outro autor faz uma análise desta modalidade, porém em forma de notas de rodapé ou como uma observação, não levando em consideração as transformações pelas quais o português tem passado.

Na análise das GTs, pudemos notar que elas apresentam alguns pontos que se mostram incoerentes e indefinidos. Dentre eles:

- a) os critérios de análise;
- b) a delimitação dos recursos de indeterminação;
- c) o contexto;
- d) os conceitos de indeterminação, indefinição e impessoalização.

3.1.1.1 Os critérios de análise

O primeiro ponto de incoerência que observamos na abordagem feita pela GT diz respeito aos critérios adotados pelos autores na conceituação do sujeito indeterminado e na apresentação dos recursos. De acordo com Perini (2004, p. 38),

as unidades lingüísticas apresentam dois aspectos fundamentais: a forma (ou “significante”) e o significado. A forma e o significado não são propriamente partes ou componentes das unidades lingüísticas, mas antes aspectos evidenciados pelo ponto de vista adotado ao se estudar a língua.

Na análise de Perini, os aspectos morfológicos e sintáticos dizem respeito ao ponto de vista formal; já o aspecto semântico diz respeito ao significado. De acordo com o autor, é

¹³ Kury (1989) especifica em nota de rodapé na p. 27.

possível, ao analisar um fenômeno lingüístico, descrever tanto os aspectos formais quanto os semânticos, separadamente. Pode-se também, com igual importância, colocá-los em correlação.

Nos compêndios gramaticais analisados, observamos que o sujeito indeterminado é apresentado juntamente com os demais ‘tipos de sujeito’: simples, composto, oculto, inexistente. Porém, ao analisarmos as definições apresentadas para esses tipos de sujeito, vemos que não há um critério entre os autores para classificá-los, visto que o sujeito indeterminado é definido a partir de um critério semântico (aquele que não é possível identificar) e os demais são definidos conforme o critério morfossintático (aquele que tem um ou mais de um núcleo; aquele que não está expresso, mas pode ser identificado pela desinência verbal etc.).

Na classificação do sujeito, apenas Rocha Lima (1980, p. 235) parte da noção de determinação para definir indeterminação:

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.

Ainda encontramos nas GTs uma confusão nos critérios de reconhecimento de um sujeito como indeterminado. Apesar de as definições apresentadas partirem de um critério semântico, há, em contrapartida, imposições de ordem puramente sintática.

Ao apresentarem o sujeito indeterminado, os autores atribuem à indeterminação o fato de se desconhecer quem realiza a ação, pautando-se, assim, num critério semântico. Porém, ao abordarem as duas formas de se indeterminar o sujeito, em uma delas o verbo deve estar na terceira pessoa do plural sem especificação do tipo de verbo; já na outra, o verbo deve estar na terceira pessoa do singular mais a partícula ‘se’, devendo este verbo ser intransitivo, transitivo indireto ou de ligação (cf. CEGALLA, 1995). Assim, é a predicação verbal, fato estritamente sintático, que vai decidir a classificação do sujeito indeterminado.

Vemos então que, enquanto para utilização da terceira pessoa do plural os critérios são de base semântica, para a utilização do ‘se’ são de caráter sintático, não havendo na GT um mesmo critério na apresentação dos recursos de indeterminação do sujeito.

Ainda em relação à abordagem da partícula ‘se’ pela GT, Milanez (1982) lembra que, para o reconhecimento do sujeito indeterminado, a GT utiliza argumentos de ordem puramente sintática. Para abordar essa questão, a autora apresenta os seguintes exemplos:

(11) Fuma-se aqui.

(12) Fuma-se charuto aqui.

Segundo a referida autora, para a GT, as frases (11) e (12) têm interpretações diferentes. Em (11), o ‘se’ é tido como índice de indeterminação do sujeito. Já em (12), o ‘se’ é considerado apassivador e a sentença é entendida como: “Charuto é fumado aqui”. Nesta frase, o sujeito está expresso e determinado: “charuto”. Vemos então que, mais uma vez, é a predicação verbal que vai decidir a classificação do sujeito indeterminado, uma vez que em (11), o ‘se’ é índice de indeterminação do sujeito, por ser o verbo intransitivo, enquanto que em (12), o ‘se’ é partícula apassivadora, por ser o verbo transitivo direto. Nesse caso, a GT pauta-se apenas no critério sintático.

Se prestarmos mais atenção, veremos que a interpretação da GT não pode ser aceita, pois mesmo em sua interpretação passiva o falante intui a existência de um agente que é desconhecido e que não podemos determinar quem é.

3.1.1.2 Delimitação dos recursos de indeterminação

Vimos que, de modo geral, as GTs apresentam apenas dois recursos que podem ser utilizados pelos falantes para indeterminar o sujeito: o uso do verbo na terceira pessoa do plural e o uso do verbo na terceira pessoa do singular acrescido da partícula ‘se’. Apenas um ou outro autor, como citamos no item 3.1, apresenta outras formas para indeterminar o sujeito. O sistema lingüístico conta ainda com outros recursos de indeterminação, como as formas pronominais. Vejamos alguns exemplos com estas formas:

(13) Doc.: O tempo tá passano mais rápido?

Inf.: Tá, pois se **nós** tamo no fim da era... (P, 12, 172)

(14) Doc.: Como é que assiste o filme e não sabe a história dela... do filme?

Inf.: Na hora que nós tá assistino assim, **a gente** sabe, mas, depois que passa, **a gente** às vez esquece. (B, 7, 78)

(15) Doc.: A feira daqui já foi boa?

Inf.: Já foi boa... antigamente os depóso aí era chei de mamona, licuri. Naquele tempo que os depóso enchia, que **eles** botava a balança aí no mei da rua pra comprar, aí no mei da rua aí ó. (P, 5, 112)

(16) Doc.: É construída mesmo da natureza, né?

Inf.: É construída mesmo na natureza. É bonito. Só tem peda lá no... no morro do Senhor Jesus. Pra cá, **você** não vê nenhuma peda. Só aquela pranada... (B, 10, 169)

Nesses exemplos, os informantes indeterminam o sujeito de seus discursos através do uso dos pronomes *nós*, *a gente*, *eles*, *você*. Em construções deste tipo, não é possível identificarmos o referente, podendo este envolver qualquer uma das três pessoas do discurso ou as três ao mesmo tempo.

Cavalcante (1998) atribui a não admissão de outros recursos pela GT como resultado da indeterminação ser vista pela gramática como um fenômeno categórico, não tendo espaço para a idéia de gradação¹⁴. De fato, os estudos sobre o fenômeno da indeterminação do sujeito no português (cf. MILANEZ, 1982; ALMEIDA, 1992; CUNHA, 1993; MENON, 1994; SETTI, 1997; CAVALCANTE, 1998; CAVALCANTE, 1999; GODOY, 1999) têm revelado uma tendência ao uso de diversas formas, além daquelas citadas pela GT, entre elas: *a gente*, *eu*, *nós*, *você*, *vocês*, *eles*, $\emptyset+V3PS$, $\emptyset+VPASSINT$, $\emptyset+VPSA$, $\emptyset+VINFL$, podendo-se ainda considerar formas como *o cara*, *o indivíduo*, *as pessoas*, *o pessoal* etc., como veremos ainda neste capítulo.

3.1.1.3 O contexto

Para que possamos interpretar os recursos de indeterminação, é de fundamental importância a consideração do contexto em que a forma está inserida. O que vemos na GT é que ela tem seu campo de análise limitado à frase, sem levar em conta o contexto. É por isso que os autores não admitem formas como *você*, *a gente*, *eles* etc. como recursos utilizados pelos falantes para indeterminar o sujeito, pois estes recursos dependem do contexto pra que sejam interpretados como tal.

3.1.1.4 Os conceitos de indefinição e de impessoalização

A partir da análise das GTs, constatamos que há uma certa confusão entre os conceitos de indeterminação, indefinição e impessoalização, chegando mesmo a confundi-los, usando a

¹⁴ Falaremos no item 3.3.3 sobre os graus de indeterminação.

mesma denominação para casos distintos. Assim sendo, é necessário esclarecermos os conceitos de indefinição e de impessoalização e como estes serão tratados, uma vez que encontramos ocorrências dos três fenômenos nos nossos *corpora*.

Kury (1989, p. 25) define sujeito indeterminado como aquele “que ocorre quando não se quer, ou não se sabe, ou não se pode expressar o sujeito de uma oração”. Ao colocar os exemplos, o autor faz uma comparação como se estes fossem equivalentes a pronomes indefinidos, como ‘alguém’, ‘qualquer pessoa’, por ter um sentido vago e por nada esclarecer quanto à identidade do agente. Para ele, o que diferencia um sujeito indeterminado de um pronome indefinido é a expressão desse sujeito. Se expreso, é pronome indefinido e, portanto, determinado, se não é expreso, é sujeito indeterminado, mesmo que ele seja apresentado semanticamente. Outros autores, como Melo (1967), consideram como sujeito indeterminado aquele “materialmente constituído por pronomes indefinidos que nada esclarecem quanto à identidade do agente”.

Na verdade, o que percebemos é que enquanto autores discutem a existência da equivalência entre indeterminação e indefinição há aqueles que, prevendo as dificuldades da questão, preferem apenas citar os recursos sem discuti-los, apresentando definições apoiadas nos critérios semânticos.

A indefinição é um fenômeno que faz parte do domínio da determinação que muitas vezes é confundido com a indeterminação, pois, de acordo com a análise que fizemos nas GTs, esses fenômenos não apresentam fronteiras bem definidas, já que ambos apresentam um conteúdo vago, impreciso. Na indefinição, a referência diz respeito exclusivamente a formas lexicais de terceira pessoa: *tudo, nada, alguém, ninguém* etc., enquanto na indeterminação a generalização é essencial, podendo envolver qualquer pessoa (primeira, segunda e terceira) de forma isolada ou simultânea.

De acordo com Milanez (1982), há alguns aspectos que devem ser levados em conta na distinção desses dois conceitos, os quais apresentaremos aqui sumariamente:

1. a indeterminação restringe-se apenas a elementos humanos, enquanto que a indefinição pode estender-se a elementos não humanos;
2. a indefinição pressupõe um conjunto fechado de elementos que remetem à 3ª pessoa e somente a ela, e os de indeterminação podem envolver qualquer pessoa.

3. na indeterminação, a generalização é essencial, já na indefinição, ela é uma possibilidade, podendo remeter tanto a uma referência individualizada como também genérica, porém limitada a um conjunto fechado.
4. as distinções morfológicas de singular e plural e o uso dos artigos o/um, por exemplo, são irrelevantes na indeterminação, pois não há alteração no conteúdo indeterminador, enquanto que estas distinções são muito importantes na indefinição, porque acentuam a idéia de quantidade.
5. na indeterminação os recursos são dependentes do contexto para a sua interpretação, já na indefinição os recursos independem do contexto, sendo analisados dentro dos limites da frase.
6. a indeterminação refere-se de modo indistinto às três pessoas do discurso; a indefinição indica de modo preciso as pessoas do discurso envolvidas, por isso é um caso de determinação.

Com relação à impessoalização, Milanez (1982, p. 48), resumidamente, a distingue da indeterminação:

a impessoalização é uma propriedade de certos verbos que, por sua própria natureza, não atribuem a seus SN-sujeitos nenhuma função referencial. A indeterminação, ao contrário, sempre envolve uma referência. Justamente é esta ausência que diferencia as construções impessoais de todas as outras formas verbais da língua.

Vejamos os exemplos abaixo:

(17) Inf.: Porque na cidade **você** só vê prédio, trinta e cinco, quarenta, quarenta e cinco... (P, 7, 139)

(18) Inf.: **Chove** muito em São Paulo. (P, 7, 157)

No exemplo (17), “você só vê prédio” apresenta o sujeito genérico *você* cuja referência não está expressa, podendo ser qualquer pessoa, mas existe. Neste caso, temos um exemplo de indeterminação. Já em (18), “Chove muito”, não há referência alguma a algo ou alguém, pois nela não há sujeito, sendo essa oração tida como de sujeito inexistente¹⁵.

¹⁵ Na GT, dentre os tipos de sujeito, há a **oração sem sujeito**, que, segundo Bechara (1992, p. 201), são aquelas que “encerram apenas a declaração contida no predicado, sem que se cogite de atribuí-la a nenhum sujeito. Em

3.2 TRABALHOS SOBRE A INDETERMINAÇÃO EM PORTUGUÊS

As pesquisas que tratam da indeterminação têm mostrado que existem várias formas para indeterminar o sujeito em português; no entanto, vimos que na GT este entendimento apresenta-se de maneira fragmentada. Em trabalhos recentes, os recursos de indeterminação têm sido analisados sob diferentes concepções teóricas em amostras de fala de diferentes regiões do Brasil. A partir dos resultados desses estudos, é possível afirmar que a indeterminação acontece não só com formas consagradas pela GT, mas também por inúmeras formas que assumem um valor indeterminado.

Vejamos, pois, brevemente, os resultados de alguns trabalhos que analisam o fenômeno da indeterminação do sujeito tanto na língua oral quanto na língua escrita. Incidiremos nossa atenção em alguns aspectos que perpassam todas as obras sumarizadas, a saber: o conceito de indeterminação, os recursos encontrados pelos autores para indeterminar o sujeito, a noção de gradação e os resultados da análise.

3.2.1 A indeterminação na língua oral

Encontramos dois trabalhos que analisam amostras de fala do estado de São Paulo: o de Milanez e o de Menon. Milanez (1982) apresenta uma descrição dos recursos de indeterminação do sujeito encontrados no *corpus* pertencente ao Projeto NURC¹⁶ de São Paulo, tendo como teoria de análise a Gramática Gerativo Transformacional, valendo-se também da Semântica e da Pragmática, uma vez que o processo de indeterminação escapa do nível puramente sintático. A autora apresenta um universo bastante amplo de recursos de indeterminação do sujeito encontrados na amostra: *a gente, nós, você, eles, eu, o/um cara, o/um indivíduo, o/um sujeito, a/uma pessoa, Ø+V3PS, Ø+V3PP, Ø+V+SE e Ø+VINFL*. Ela ainda ressalta que o inventário dos recursos de indeterminação pode ser ainda mais amplo.

tais casos dizemos que se trata de *orações sem sujeito* e o verbo que nelas entra se chama *impessoal*". Os principais verbos impessoais são: a) os que denotam fenômeno da natureza; b) o verbo *haver* nas orações que denotam a existência de pessoa ou coisa; c) o verbo *ser* nas indicações de tempo.

¹⁶ O projeto NURC (Projeto de Estudo da Norma Lingüística Urbana Culta no Brasil) vem-se desenvolvendo desde 1969 em cinco capitais brasileiras: Salvador, Recife, Rio de Janeiro, São Paulo e Porto Alegre. Tem como objetivo descrever os padrões reais de uso na comunicação oral adotadas por falantes de nível superior em cada uma das cidades citadas. Os informantes foram distribuídos pelos dois sexos e por três faixas etárias. Os discursos que compõem os *corpora* distribuem-se em três categorias de texto: elocuições formais (EF), diálogo entre informante e documentador (DID) e diálogo entre dois informantes (D2).

Ao descrever os recursos encontrados no *corpus*, Milanez divide-os em dois grandes grupos: as formas sem sujeito lexical e as formas representadas por sujeito lexical, procurando mostrar as características sintático-semânticas de cada uma e a forma como ocorreram na amostra. Segundo a autora, “a indeterminação não se apresenta como um fenômeno de natureza absoluta, que se opõe à determinação sempre de forma nítida; ao contrário, a mesma comporta ‘graus’ quanto à sua abrangência” (p. 80).

Para Milanez (1982), as formas sem sujeito lexical, $\emptyset+V3PS$, $\emptyset+V+SE$ e $\emptyset+VINFL$, são as que expressam o mais alto grau de indeterminação, ou seja, apresentam indeterminação completa. Isso porque não são marcadas quanto à pessoa, distanciando-se de referências específicas a qualquer das três pessoas do discurso. As outras formas com sujeito pronominalizado, a depender do pronome, apresentam um grau de indeterminação intermediário, pois, embora envolva pessoas não determinadas pelo contexto, mantêm referências específicas aos interlocutores do diálogo. *A gente* e *eu*, por exemplo, incluem obrigatoriamente a primeira pessoa, enquanto *você* sempre envolve a segunda pessoa. O menor grau de indeterminação é expresso pela forma $\emptyset+V3PP$ e pelo pronome *eles*, pois são recursos exclusivos de terceira pessoa, além de ocorrerem normalmente antecidos de uma referência a um grupo social a que pertence o agente verbal.

Assim, segundo Milanez, dois fatores são relevantes para identificação dos graus de indeterminação: a própria forma lexical do recurso e o contexto. O primeiro porque de certa forma já determina a amplitude da generalização dos recursos e o segundo por desempenhar um papel de controle dessa abrangência.

Em seu trabalho, Milanez analisa as formas despronominalizadas de indeterminação à luz da Gramática Gerativo Transformacional. A partir da Teoria Padrão Estendida, ela busca fundamentação para não considerar a forma $\emptyset+V3PS$ como o resultado de apagamento de algum pronome pessoal usado com verbo na 3ª pessoa do singular nem de apagamento de uma construção de sentido equivalente ao pronome indefinido *alguém*. Para ela, tanto essa forma quanto as demais formas despronominalizadas constituem casos em que a referência é arbitrária e não anafórica.

Milanez conclui o seu trabalho dizendo que há uma grande distância entre o que é ensinado nas escolas e o que é característico da fala, mesmo no dialeto culto. Conclui ainda que a indeterminação é um processo da língua explicável essencialmente a nível semântico-pragmático e, por isso, se revela um fenômeno dependente de contexto para sua interpretação.

É importante ressaltar que o trabalho de Milanez representa uma grande contribuição ao estudo do português oral do Brasil, além de ter servido de referência para diversos trabalhos já realizados sobre a indeterminação do sujeito.

A partir da análise de 68 entrevistas que compõem o *corpus* do Projeto NURC de São Paulo, há também o trabalho de Menon (1994), que faz uma abordagem sociolinguística do fenômeno da indeterminação a partir da análise de três variáveis sociais: sexo, idade e estilo. Menon analisa 12 formas que o falante culto paulistano utiliza para indeterminar o sujeito em seus enunciados: *a gente, eles, eu, Formas Nominais (FNs), nós, se, você, vocês, VPSA, VPASSINT, Ø+V3PP, Ø+V3PS*.

A primeira hipótese defendida por Menon refere-se aos meios pelos quais o sujeito indeterminado pode ser expresso no português falado no Brasil. Segundo ela, haveria mais meios do que os previstos pela GT, o que é comprovado em sua análise, pois ela encontrou dez recursos que o falante utiliza para indeterminar o sujeito em português, além dos dois previstos pela GT.

A segunda hipótese refere-se às variantes linguísticas da indeterminação. A partir de quatro delas, a autora chega às seguintes conclusões: 1) a forma *a gente* estaria passando por um processo de gramaticalização, tornando-se um pronome, a princípio de caráter genérico, em seguida indeterminado e, mais recentemente, utilizado no lugar das primeiras pessoas, sobretudo do plural; 2) para preencher o vazio deixado pela forma *a gente*, os falantes estariam utilizando a forma *a pessoa*; 3) o pronome *se* é um sujeito de pleno direito; 4) *você* transformou-se em pronome pessoal de segunda pessoa, em concorrência com *tu*, e é também um recurso de indeterminação do sujeito.

A terceira hipótese refere-se à idade, ao sexo e ao estilo. A autora afirma que o uso das diferentes formas de indeterminação do sujeito variam de acordo com a faixa etária do informante. De acordo com seus resultados, as formas *eu, eles* e *você* são “inovadoras” por serem utilizadas com maior frequência e de modo regular pelos mais jovens. As formas *Ø+V3PP, nós* e *VPASSINT* são mais usadas pelos falantes da terceira faixa etária, sendo, portanto, arcaizantes. E a forma *Ø+V3PS* está estabilizada na língua já há algum tempo, pois é utilizada pelas três faixas etárias.

Com relação ao sexo, Menon quer verificar se em seus dados as mulheres são mais conservadoras. Em relação aos homens, ela investiga se são eles que tendem a inovações. De acordo com os resultados, são as mulheres que utilizam ou veiculam as formas novas, enquanto os homens utilizam os recursos mais formais de indeterminação.

Quanto ao estilo, a autora refere-se ao tipo de entrevista. Sua hipótese é a de que as formas podem se especializar conforme os tipos de texto e de situação. Menon esperava que nas elocuições formais (EFs) as formas consideradas pela GT seriam mais numerosas. Por outro lado, ela esperava encontrar uma utilização maior de FNs nos diálogos entre informante e documentador (DIDs) e nos diálogos entre dois informantes (D2), devido ao seu caráter mais informal em que se utilizam formas características da oralidade. A análise dos dados mostrou que realmente houve uma variação de formas conforme o estilo, mas que também há formas que transitam pelos três estilos, sendo elas fortes candidatas a se tornarem estáveis na língua.

Ainda há os trabalhos de Almeida (1992) e Cunha (1993) que analisam amostras de fala de informantes residentes na cidade do Rio de Janeiro. Almeida (1992) analisa a indeterminação do sujeito a partir das entrevistas do acervo do Projeto Censo da Variação Lingüística¹⁷, com informantes de ambos os sexos de nível básico e médio de escolaridade. Neste trabalho, a autora estuda a indeterminação do sujeito sob dois aspectos: a indefinição e a referencialidade. Na análise dos dados, ela parte dos pressupostos teórico-metodológicos do Funcionalismo de base givoniana e da Sociolingüística quantitativa.

Segundo Almeida, a indeterminação do sujeito consiste na vaguidade referencial, ou seja, o referente não é definido. Ela ainda ressalta que quando a indeterminação é realizada no fluxo discursivo, ela ocorre com diferentes graus de indefinição referencial e é codificada por diferentes formas, como as ditas “camaleônicas”, quais sejam *os caras, o pessoal, as pessoas*, entre outras, além das formas pronominais *eu, a gente, nós, tu, você, eles*, expressas ou não, e o clítico *se*.

A autora observa que, em relação aos graus de indefinição do referente, o sujeito pode ser introduzido pelo falante de modo determinado e ir se indeterminando ou vice-versa, delimitando três graus de acordo com a sua indefinidade.

¹⁷ O Projeto Censo é constituído por um banco de dados organizado por uma equipe de pesquisadores sediada na Faculdade de Letras da UFRJ, congregando também professores de outras universidades - a equipe do Projeto Censo. É um grupo de pesquisa interinstitucional que vem sendo desenvolvido desde o final dos anos 1970, sob organização de Anthony Julius Naro. As pesquisas são voltadas para o estudo do binômio variação e mudança e da interface língua e sociedade. O grupo de pesquisadores se congregou inicialmente em torno do objetivo de analisar fenômenos lingüísticos variáveis e depreender os possíveis processos de mudança em curso na variedade não culta falada na cidade do Rio de Janeiro. Para a consecução deste objetivo o grupo empreendeu a formação de uma amostra da fala carioca que ficou conhecida como Amostra Censo. Ainda perseguindo a compreensão da dinamicidade do português, o grupo se voltou mais recentemente para um estudo da mudança em tempo real de curta duração, realizando pesquisas do tipo painel (centrada na mudança no comportamento lingüístico do indivíduo) e tendência (centrada na comunidade) através da constituição de novas amostras de fala dos informantes que compõem a Amostra Censo e da comunidade de fala carioca.

Cunha (1993), partindo das indicações de Milanez (1982), analisa 18 gravações do acervo sonoro do projeto NURC do Rio de Janeiro e tem como objetivo delimitar quais as condições lingüísticas e sociais que levam à escolha de *nós*, *você* e *a gente* como indeterminadores do sujeito. Para isso, ela analisa a amostra através de uma abordagem teórica da Análise do Discurso e da Sociolingüística.

A autora faz uma análise qualitativa e uma análise quantitativa dos dados. Na análise qualitativa, foi possível fazer três afirmações. A primeira delas foi em relação ao conceito de indeterminação. Para ela, “um pronome para ser considerado indeterminado não deve ter sua referência diretamente recuperável, podendo ser associado a um leque de referentes depreensíveis do contexto” (p. 43).

A segunda afirmação foi a de que a indeterminação manifesta-se em maior ou menor grau a partir dos pronomes e do seu uso dentro de um contexto. Assim, a autora distingue três graus de acordo com os diferentes níveis de referência no contexto: 1) indeterminação completa, que ocorre quando não há elementos no contexto para indicar a referência; 2) indeterminação parcial com referência implícita no contexto, quando a referência é depreendida do contexto sem que haja um item lexical presente no texto que sirva de guia para a interpretação; e 3) indeterminação parcial com referência explícita no contexto, quando a referência é depreendida do contexto a partir de itens lexicais presentes no texto, o qual permite a sua interpretação.

A terceira afirmação da autora visa estabelecer “entre o contexto e a indeterminação pronominal uma relação de dependência: quanto mais a interpretação do pronome depende de elementos contextuais, mais próximo ele está de ser indeterminado” (p. 43-4).

Na análise quantitativa, a partir dos pressupostos metodológicos da Sociolingüística Quantitativa Laboviana, a autora testou variáveis independentes como: tempo verbal, tipo de oração, tipo de elocução, gênero discursivo, graus de indeterminação, tamanho do grupo, mudança de referente, forma antecedente, sexo e faixa etária e chegou às seguintes afirmações:

- 1) Confirma-se a hipótese de que fatores intra e extra lingüísticos influenciam o processo de indeterminação pronominal.
- 2) Os fatores que atuam na escolha do pronome são, primordialmente, os de natureza discursiva, sendo que os fatores sociais ocupam uma posição secundária e os fatores estritamente gramaticais não se mostram relevantes.

- 3) Uma vez escolhida a forma é acionado um mecanismo de repetição, isto é, repete-se uma ou mais vezes a forma escolhida até que o contexto determine o uso de outra. (p. 102)

Ao concluir seu trabalho, a autora mostra que a indeterminação pronominal, especialmente quando expressa pelas formas *nós*, *você* e *a gente*, é um fenômeno sensível a condicionamentos lingüísticos e sociais. Além disso, ela percebeu que a variação *nós* e *a gente* também se dá quando os pronomes são indeterminados, havendo, porém, contextos que propiciam a escolha de uma e não de outra forma.

Rolleberg *et al* (1991), através de um estudo-piloto, observam a indeterminação nos pronomes pessoais *você*, *nós*, *a gente* e *eles* no *corpus* do projeto NURC de Salvador e os relaciona com duas variáveis sociolingüísticas: categoria de texto e faixa etária.

As autoras consideram a indeterminação do sujeito como indeterminação da referência do sujeito. Para elas, o sujeito é indeterminado quando não é possível, dentro de um contexto discursivo, especificar nominal ou numericamente sua identidade. Esta afirmação nos mostra que a referência indeterminada não pode ser recuperada, pelo fato de não ter sido precisada ou determinada na totalidade do discurso. A indeterminação para elas se opera através de recursos que exprimem generalização, porém alguns recursos são mais generalizadores que outros e, assim como Milanez (1982), elas atribuem essa amplitude da generalização à própria forma de expressão e ao contexto em que ocorre o recurso indeterminador. Desta forma, o conteúdo de um recurso pode variar a depender das condições contextuais em que figure.

Ao final, as autoras propõem a escritura de uma gramática descritiva do português falado, o que já vem sendo feito por alguns lingüistas, atestando a distância que há entre as línguas falada e escrita.

Cavalcante (1998) apresenta uma descrição dos recursos de indeterminação do sujeito encontrados no português oral culto de Fortaleza – PORCUFORT –, analisando um *corpus* com 35 informantes. Em seu estudo, a autora considera os seguintes recursos de indeterminação, que divide em três grupos¹⁸: G1 (*eu, você, a gente, nós*); G2 (\emptyset +verbo na 3ª p. pl[*eles*], substantivos genéricos [*a/uma pessoa, o/um sujeito, o/um pessoal, as pessoas, o/um cabra, o /um cara*]); G3 (*se + infinitivo, se+verbo na 3ª p. sg., \emptyset + 3ª p. sg.*).

O trabalho de Cavalcante constitui uma descrição do português oral a partir do ponto de vista da GT sobre o qual ela propõe críticas e avanços para o tratamento da indeterminação. Com este trabalho, a autora busca: a) mostrar as formas que caracterizam a indeterminação no

¹⁸ A divisão em grupos é feita a partir da referência às pessoas do discurso que engloba cada uma das formas de indeterminação.

português oral de Fortaleza; e b) compreender o uso dos recursos de indeterminação a partir da referência e da modalização imposta pelo emissor ao seu enunciado. Em sua análise, a autora utiliza-se da Sociolinguística e da Teoria da Enunciação.

De acordo com a autora, o conceito de indeterminação está ligado à referência genérica às três pessoas do discurso. A partir do tipo de referência que cada um dos indeterminadores proporciona, será possível considerar diferentes graus de indeterminação, pois o conteúdo genérico de cada recurso tem características diferenciadas. Para ela, “a elaboração de uma escala só pode ser feita em função do que a forma lexical dos indeterminadores indicar: inclusão ou exclusão das pessoas do discurso” (p. 87)

Os resultados da análise mostram que os indeterminadores do G1 possibilitam em seu uso tanto afastamento como aproximação em decorrência das várias atitudes tomadas pelo falante em relação ao seu enunciado. Assim, a autora conclui que, de modo geral, é a fusão entre a referência genérica às pessoas do discurso e a modalização imposta pelo falante ao enunciado que determina a opção do falante por um uso indeterminado.

Setti (1997) estuda algumas possibilidades de indeterminação do sujeito no discurso oral, a partir da análise de 62 entrevistas feitas nas três capitais do Sul do Brasil (Curitiba, Florianópolis e Porto Alegre) que compõem o banco de dados do Projeto VARSUL¹⁹. Foram selecionadas as seguintes possibilidades de indeterminação do sujeito: *a gente, eles, eu, formas nominais* (como *o camarada, o indivíduo, a pessoa*), *nós, se, tu, você(s), voz passiva sem agente, voz passiva sintética, Ø V3PS e Ø V3PP*. Para análise dos dados, a autora segue o modelo variacionista proposto por Labov, selecionando as variáveis tempo e modo verbal, faixa etária, sexo e localidade.

De acordo com Setti, o sujeito é indeterminado “quando não há possibilidade ou não há interesse em se identificar o seu referente num dado contexto” (p. 2). Para ela, baseando-se em Menon (1994), alguns critérios condicionam a indeterminação: a intercambialidade das formas; os pares mínimos; os ditados, verdades gerais ou eternas, questões retóricas; tempos e modos verbais; advérbios e preposições; completivas; afastamento do falante; afastamento no tempo, situações hipotéticas. A autora não faz referência à idéia de gradação, trabalhando

¹⁹ O projeto VARSUL vem-se desenvolvendo em algumas universidades da região Sul do Brasil, UFPR, UFSC, UFRGS e PUC-RS, com a finalidade de trabalhar com as variações linguísticas do sul. É formado por um banco de dados com as gravações de entrevistas feitas com 24 informantes de cada cidade do Sul, sendo as 3 capitais (Curitiba-PR, Florianópolis-SC e Porto Alegre-RS) mais 3 cidades de cada estado. Foram levados em consideração, na montagem da amostra, o sexo dos informantes, a idade (25 a 50 anos e mais de 50 anos) e o grau de escolaridade (primário, ginásio e 2º grau).

apenas com recursos que apresentam caráter genérico total, excluindo os casos de indeterminação parcial.

A autora chega à conclusão de que, apesar das diferenças geográficas, etárias, culturais e outras englobadas em cada estado, os falantes têm em comum o contato com um universo de recursos lingüísticos amplo. Em seu estudo, ela comprova que há um uso efetivo de outras variantes para indeterminar o sujeito e não apenas as citadas pela GT.

Godoy (1998) descreve o fenômeno a partir de uma amostra do português falado em Irati, Londrina e Pato Branco, cidades do interior do Paraná pesquisadas pelo projeto VARSUL, obtidas em 72 entrevistas do tipo DID. Ela apresenta como formas de indeterminação as variantes *a gente*, *eles*, *eu*, *formas nominais (o camarada, o indivíduo, a pessoa)*, *nós*, *se*, *tu*, *você(s)*, *voz passiva sem agente* e \emptyset V3PP. Os dados são analisados a partir da metodologia variacionista laboviana, selecionando como grupos de fatores as mesmas variáveis selecionadas por Setti (1997).

Para a autora, o sujeito é indeterminado quando não há possibilidade de identificação do referente no contexto em que ele se insere. A autora, assim como Setti (1997), não leva em consideração os graus de indeterminação. Apresenta apenas, também baseada em Menon (1994), alguns condicionamentos lingüísticos na seleção dos dados: a intercambialidade das formas, os pares mínimos, os tempos verbais, os advérbios, as preposições, as completivas, os recursos que marcam o distanciamento do falante e o afastamento no tempo e o discurso hipotético.

O estudo de Lopes (1993), apesar de não se apresentar como um trabalho sobre a indeterminação, constitui uma abordagem variacionista das formas *nós* e *a gente* no português falado culto. A amostra analisada é composta de dados distribuídos por três regiões do país: Sudeste (Rio de Janeiro), Sul (Porto Alegre) e Nordeste (Salvador), que compõem o *corpus* do Projeto NURC.

Um dos grupos de fatores considerado pela autora é o ‘eu-ampliado’. Este é considerado a partir de níveis de abrangência, situado entre dois extremos: grau máximo de inclusão do ‘eu’ e grau mínimo de inclusão do ‘eu’. Considera, então, três níveis em que as formas estudadas (*nós* e *a gente*) podem ocorrer: 1) o aspecto inclusivo [+ determinado], ‘eu’ + ‘não-eu’; 2) o aspecto exclusivo [+ ou – determinado], ‘eu’ + ‘não-pessoa’; e 3) o aspecto genérico [- determinado], ‘eu’ + ‘não-eu’ + ‘não-pessoa’.

Analisamos também o trabalho de Machado (1995) que busca averiguar os condicionamentos estruturais e sociais da alternância de emprego de *nós* e *a gente* em variantes populares não-urbanas na região Norte do estado do Rio de Janeiro, a partir dos

dados do Atlas Etnolingüístico dos Pescadores do Estado do Rio de Janeiro (APERJ). Nesse trabalho, a autora considera também como um de seus grupos de fatores ‘o grau de determinação do referente’. A classificação gradual de determinação do referente utilizada pela autora foi definida em função da noção de ‘eu-ampliado’.

Assim, considera três níveis: 1) casos de referente determinado ou explícito, que incluem ‘eu’ + ‘não-eu’, ‘eu’ + ‘não-eu’ com referência explícita e ‘eu’ + ‘não-eu’ + ‘não-pessoa’ com referência explícita; 2) casos de referente parcialmente determinado, que incluem ‘eu’ + ‘não-pessoa’ não determinada com elo de referência explícito e ‘eu’ + ‘não-eu’ + ‘não-pessoa’ não determinada com elo de referência implícito; e 3) casos de referente [-determinado], que agrupam as ocorrências de ‘eu’ + outras pessoas totalmente indeterminadas.

3.2.2 A indeterminação na língua escrita

Embora os estudos tenham se concentrado na língua oral, a língua escrita também tem sido estudada no que se refere à indeterminação. Tupiná (1984) analisa um *corpus* constituído por contos brasileiros contemporâneos e por ocorrências colhidas no jornal Folha de São Paulo e na revista ISTO É.

A autora aborda a indeterminação do agente considerando o fenômeno a partir da referência pessoal. Para ela, “a indeterminação corresponde ao caráter de indiferenciação, falta de individualidade ou de especificidade de um termo, capaz de conferir ao enunciado um teor de imprecisão do ponto de vista do emissor” (p. 63). Considera a indeterminação como um fenômeno ligado ao enunciado, pois o sujeito da enunciação recorre a ela quando não pode ou não quer precisar sua idéia, sentimento ou situação. Assim, elenca os seguintes processos de indeterminação do agente: *verbo na 3ª pessoa do plural, verbo na 1ª ou na 2ª pessoa do singular ou do plural, pronomes indefinidos, você, infinitivo, voz passiva analítica, voz passiva pronominal e ‘se’ junto a verbo na 3ª pessoa do singular*.

No que se refere à gradualidade dos indeterminadores, a autora considera que os pronomes, às vezes, apresentam uma gradação, “passando da generalização total, com um recurso de maior abrangência, ainda que particularize a inclusão, para um de menor abrangência, ainda que particularize a inclusão de uma das pessoas do discurso” (p.68).

Cavalcante (1999) faz um trabalho de cunho diacrônico que tem como objetivo observar diferentes estratégias de indeterminação (formas pronominais e usos do clítico *se* em sentenças finitas e não-finitas) na escrita de jornais cariocas dos séculos XIX e XX. Essa

observação teve como intuito tentar capturar uma possível mudança nessa representação e relacioná-la a outros fenômenos de mudança em curso no PB, associando os pressupostos teórico-metodológicos do Modelo de Princípios e Parâmetros da Teoria Gerativa e da Sociolingüística Variacionista. Além disso, ela busca avaliar se as prescrições ditadas pela GT para a norma padrão culta conferem com a norma padrão dos jornais.

A autora constatou que há realmente uma gama maior de estratégias de indeterminação em sentenças finitas do que as apresentadas pelas GT e que existem reflexos das mudanças da língua oral na língua escrita padrão, inclusive para os sujeitos de referência arbitrária, consagrando, assim, novas formas de indeterminação na modalidade escrita da língua.

3.3 A INDETERMINAÇÃO SOB A NOSSA PERSPECTIVA

A partir dos conceitos apresentados nessas pesquisas, vimos que definir o que seja um sujeito indeterminado não é uma tarefa fácil, visto que, utilizando as palavras de Milanez (1982, p. 23), “a indeterminação é um fenômeno complexo que dá margens a estudos extensivos em diferentes níveis”.

Consideremos o trecho abaixo retirado dos nossos *corpora*:

- (19) Doc.: Lá é muito diferente daqui, né? Aqui é tranqüilo. Você sentiu assim, aquele baque, não, quando chegou lá?
 Inf.: Ah! Logo no início **a gente** chega assim na primeira viagem **a gente** sente, né? Mai logo... pelo meno **a gente** costuma.
 Doc.: É, com o tempo, né?
 Inf.: É, com o tempo **a gente** costuma, mai a primeira, a primeira viagem que **você** vai, a primeira vez, aí sente sim, porque aqui é um lugar calmo, lugar sem agitação, quando **a gente** chega lá, num lugar daquele lá todo, todo lugar, todo agitado, os pessoá só anda o máximo às vez na carreira. (B, 9, 130).

Nesse exemplo, observamos na fala do informante a existência do sujeito sintático expresso por *a gente* e *você*, porém nenhum deles se refere aos participantes do discurso, sendo uma referência indeterminada, uma vez que o contexto não permite que digamos exatamente ‘quem chega a algum lugar’, ‘quem sente algo’ ou ‘quem vai a algum lugar’, podendo referir-se a qualquer pessoa. Vemos que estas formas, neste contexto, têm um conteúdo generalizado, englobando as três pessoas do discurso.

Para a análise do conceito de indeterminação, Milanez diz que é necessário que se leve em conta noções como a de pessoa, a de generalização e a de referência. Para Cavalcante

(1998), a noção de referência se destaca, pois é através dela que a noção de pessoa e de generalização se organiza. Concordamos com as referidas autoras e é a partir dessas noções que procuraremos delimitar o conceito de indeterminação.

3.3.1 Determinação e indeterminação

Inicialmente é importante distinguir a determinação da indeterminação. De acordo com Cavalcante (1998), é a partir da referência, apresentando-se específica ou não, que se instaura a determinação e a indeterminação. Para ela,

a determinação da referência às pessoas do discurso define-se, então, na medida em que aponta para as pessoas do discurso, especificando-as; enquanto a referência indeterminada é entendida a partir do pressuposto de que não é possível apontar, com precisão, as pessoas envolvidas no discurso, pois apresenta-se sob a forma genérica, podendo abranger todas as pessoas do discurso. (p. 34)

Assim, observamos que, como dizem Milanez (1982) e Cavalcante (1998), a indeterminação se opera através de recursos que exprimem generalização. Sendo assim, a abrangência desses recursos pode atingir, indistintamente, qualquer uma das pessoas do discurso ou as três ao mesmo tempo, sendo que é obrigatório o envolvimento da ‘não-pessoa’.

Vejamos os trechos abaixo:

(20) Inf.: Minha mãe morreu no ano de quinze de uma melencia. Trabeiava mais ela, e ela tava de gravidez e foi comer uma melencia que **nós** achemo dento da mandioca veia. Eu digo “mãe, num come essa melencia, a senhora vai comer essa melencia, essa melencia lhe faz má”. Eu toda vida fui ativo. (P, 13, 168)

(21) Doc.: O senhor nasceu aqui?

Inf.: Não.

Doc.: Nasceu onde?

Inf.: Na Feira de Santana.

Doc.: Nasceu lá? Eu moro lá.

Inf.: **Você** mora na Feira de Santana? (P, 13, 165)

(22) Doc.: Como é que faz aquilo?

Inf.: Ali é... **Ø cavaca** a terra, **Ø móia**, ma... **Ø móia** bem moiadim e **Ø massa** o barro bem massadinho mó dele ficar assim liguento. Aí, **Ø** põe na fôrma e **Ø soca** tudo e forma o adobo. **Ø Chama** adobão. (B, 7, 79-80)

Ao analisar esses trechos, vemos que em (20), o referente do pronome *nós* está claramente especificado no contexto: o locutor e a mãe dele. Nesse exemplo, o referente não

deixa dúvida sobre a sua determinação. Em (21), o pronome *você* tem seu referente igualmente explicitado, sendo a referência direta à segunda pessoa. Neste caso, é também determinado.

Já em (22), não está claro no contexto quem é o referente, sendo este qualquer pessoa que vá fazer o adobão. Sua referência pode envolver todas as pessoas do discurso. Essa ocorrência, apesar de ser uma forma de terceira pessoa, não nos remete a um ‘ele’ que está fora da situação dialogal, mas a um nível de generalização que possibilita uma interpretação de inclusão dos interlocutores na ação verbal, sendo a referência de tal forma abrangente que pode envolver qualquer pessoa. Estamos, pois, diante de um processo de indeterminação.

3.3.2 Noção de pessoa

De acordo com Cavalcante (1998, p. 45), “a compreensão da noção de pessoa é necessária para a definição dos papéis desempenhados pelo sujeito no ato da comunicação”. Pessoa é uma categoria que se fundamenta na referência aos participantes da comunicação. A noção de pessoalidade parece ser um pouco mais complicada do que a tradição gramatical sugere. Benveniste (1995), ao descrever o funcionamento do sistema pronominal, estabelece que o eixo comunicativo comporta duas pessoas: a primeira, o ‘eu’ (locutor), e a segunda, o ‘tu’ (interlocutor). A chamada terceira pessoa é definida como membro não marcado da correlação de pessoa, nomeada pelo autor como ‘não-pessoa’.

Benveniste explica que:

nas duas primeiras pessoas, há ao mesmo tempo uma pessoa implicada e um discurso sobre essa pessoa. ‘Eu’ designa aquele que fala e implica ao mesmo tempo um enunciado sobre o “eu”: dizendo ‘eu’, não posso deixar de falar de mim. Na segunda pessoa, “tu” é necessariamente designado por ‘eu’ e não pode ser pensado fora de uma situação proposta a partir do “eu”; e, ao mesmo tempo, ‘eu’ enuncia algo como um predicado de “tu”. Da terceira pessoa, porém, um predicado é bem enunciado somente fora do “eu-tu”; essa forma é assim exepuada somente pela qual “eu” e “tu” se especificam. Daí, ser questionável a legitimidade dessa forma como “pessoa”. (p. 250).

De acordo com o que foi dito pelo autor, a primeira e a segunda pessoa do discurso possuem realmente marca de pessoa, ao passo que a terceira é privada desta marca. Benveniste (1995, p. 250) ainda afirma que “a forma dita de terceira pessoa comporta realmente uma indicação de enunciado sobre alguém ou alguma coisa, mas não referida a uma ‘pessoa’ específica”. Desta forma, pronomes de primeira e segunda pessoa tornam-se

indeterminados na medida em que assumem a capacidade de englobar a ‘não-pessoa’, o que pode acontecer inclusive com o pronome *eu*.

A noção de pessoa contida nos pronomes pessoais está intimamente ligada à noção de número. Ainda de acordo com Benveniste (1995), as duas primeiras pessoas do discurso não são passíveis de multiplicação, mas de ampliação, pois o que aumenta é a abrangência do seu conteúdo. Desta forma, o *nós* (primeira pessoa do plural) não significa ‘eu + eu’, mas ‘eu + tu’ ou ainda ‘eu’ + ‘tu’ + ‘não-pessoa’, constituindo-se um ‘eu-ampliado’. Segundo o mesmo autor, somente a terceira pessoa (não-pessoa) admite o verdadeiro plural, sua multiplicação é possível por não ser ela marcada pela unicidade de pessoas do discurso.

3.3.3 Graus de indeterminação

A indeterminação “não se apresenta como um fenômeno de natureza absoluta, que se opõe à determinação de forma nítida; ao contrário, a mesma comporta graus quanto à sua abrangência” (MILANEZ, 1982, p. 80). Como foi colocado por alguns estudiosos citados no item 3.2, os recursos usados pelo falante para indeterminar o sujeito não agem da mesma forma, o conteúdo genérico de cada recurso tem características referenciais diferenciadas. Apesar de os recursos apresentarem algumas características específicas quanto à sua referência, na sua realização eles apresentam uma capacidade de generalização que permite uma grande oscilação de seus usos indeterminados.

Alguns autores acreditam que a noção de gradação está relacionada exclusivamente à inclusão ou não inclusão das pessoas do discurso. Acreditamos, assim como Cunha (1993), que a indeterminação cresce à medida que deixam de existir elementos no contexto que permitam a interpretação do pronome. Desta forma, o grau de indeterminação será maior ou menor de acordo com a dependência da interpretação em relação ao contexto.

Em nossos *corpora* a indeterminação ocorre também de forma gradativa dentro dos limites impostos pelo contexto. Analisemos os trechos abaixo:

(23) Doc.: E como é que faz rapadura mesmo?

Inf.: Rapadura, **cê** pega a cana, passa ela no engenho, aí sai aquele cardo. Aí **você** pega, põe aquela garapa num... numa tacha e aí agora **você** põe e come. (B, 9, 143).

(24) Doc.: O sustento vocês tiram tudinho daqui, o sustendo de vocês?

Inf.: **A gente** compra assim pouca coisa. Assim arroz, arroz **a gente** compra, **a gente** não pranta arroz. Doce, açúcar, olé, esses aí... **a gente** compra, mas

feijão, milho, milho tem hora que Ø não pranta, que Ø tem o da lavora, mas farinha de mandioca, essas coisa assim não se compra não. Galinha mermo, **a gente** mata assim no quintal. (B, 4, 58)

- (25) Inf.: ... a única coisa mais longe qu'eu já fui foi na Lapa.
 Doc.: Sim. A senhora foi como?
 Inf.: **Nóis** foi de caminhão.
 Doc.: Na romaria?
 Inf.: Na romaria. (B, 8, 169)

No exemplo (23), o pronome *você* tem um conteúdo generalizado se referindo às três pessoas do discurso e significando qualquer pessoa que vá fazer rapadura. Não é possível determinar, a partir do contexto, quem, especificamente, realiza a ação.

No exemplo (24), apesar de não haver no contexto palavras que explicitem a referência do pronome *a gente*, podemos inferir que ele representa um grupo (eu, o informante, mais as pessoas que vivem na comunidade), mas isso só é depreendido através do sentido global do texto.

Já no exemplo (25), é possível depreender a referência do pronome *nós* a partir de elementos explícitos do texto. Compreendemos que, ao usar esta forma, a informante se refere a ela mesma mais um grupo de pessoas: as pessoas da comunidade que vão para a romaria, o que é possível determinar a partir de pistas que estão explícitas no contexto. Mesmo neste caso, o sujeito conserva um grau de indeterminação, pois não sabemos quem são as pessoas que vão para a romaria.

A partir desses exemplos, julgamos haver, assim como Cunha (1993), três graus de abrangência dos recursos de indeterminação que se estabelecem de acordo com as pessoas que englobam e com os diferentes níveis de referência no contexto:

- 1º Indeterminação completa
- 2º Indeterminação parcial com referência implícita
- 3º Indeterminação parcial com referência explícita

Vendo de baixo para cima, a indeterminação cresce à medida que deixam de existir elementos no contexto que permitam a interpretação da forma de indeterminação. Relacionaremos os graus de indeterminação às formas analisadas neste trabalho no capítulo da análise dos dados, visto que tomaremos o grau de indeterminação como um dos fatores lingüísticos que condicionam o uso dos recursos de indeterminação usados pelos falantes das comunidades analisadas.

É interessante ressaltar que uma mesma forma pode assumir diferentes graus de indeterminação a depender da referência e do contexto discursivo, pois o potencial generalizador de certos recursos é flexível. Vejamos como se comporta o indeterminador *a gente* nos três exemplos abaixo:

- (26) Doc.: E que música a senhora gosta?
 Inf.: [rindo]. Ixe! Eu gosto quase todas as musga.
 Doc.: Ah!
 Inf.: A gente gosta é de... Tem umas musiga que **a gente** não gosta não, né?
 Doc.: É.
 Inf.: Mas tem outas que **a gente** adora mais. (B, 12, 227)
- (27) Inf.: Cobra não. Cobra aqui sumiu com essa seca, é difícil **a gente** vê uma cobra.
 Doc.: É mermo? Mais já teve muita cobra aqui, né?
 Inf.: Já teve um bando, mas agora **a hente** num vê não, sumiu. (P, 2, 40)
- (28) Doc.: Hum. Quando você era criança, brincava de quê assim?
 Inf.: [rindo] Eu não sei não.
 Doc.: Não lembra mais? De roda?
 Inf.: É. **A gente** brincava de roda nas porta, brincava de boneca, fazer casinha.
 (B, 4, 49)

Nestes exemplos, vemos que o mesmo recurso tem referências diferentes, portanto, graus de abrangência diferenciados. No exemplo (26), há uma extensão de pessoas envolvidas na forma *a gente* utilizada pela informante. Ao falar das preferências musicais, ela generaliza, estendendo sua opinião não somente ao gosto dela, mas ao de qualquer pessoa, envolvendo ela mesma, seu interlocutor e outras pessoas. Vemos aí um maior grau de indeterminação, sendo este um exemplo de indeterminação completa da qual falamos anteriormente.

No exemplo (27), o informante utiliza o mesmo recurso para indeterminar o sujeito – *a gente* –, porém o seu uso apresenta um conteúdo um pouco diferenciado do primeiro. Ao falar da existência de cobra, o informante usa *a gente* para referir-se a ele mesmo e às pessoas da comunidade, não incluindo o documentador (interlocutor). Embora não existam no contexto palavras que explicitem esta referência, ela é depreendida a partir do sentido global do texto.

Já no exemplo (28), mais uma vez a informante utiliza a forma *a gente*, porém com um conteúdo referencial diferente dos dois anteriores. É possível inferir, a partir de palavras existentes no contexto, que a informante fala de uma época anterior à atual (a época de sua infância) e, ao usar a forma *a gente*, refere-se a ela mais às crianças de sua época, excluindo o interlocutor. Diferentemente do exemplo (27), a referência às crianças está explícita na fala do documentador o que faz com que identifiquemos a que grupo a informante está se referindo.

3.3.4 Equivalência e alternância

Um outro fator importante de ser observado é a equivalência dos recursos de indeterminação. Concordamos com Cavalcante (1998) quando diz que um recurso pode alternar, mas não equivaler exatamente ao anterior, pois “o conteúdo referencial muda de uma forma para outra e também dentro de uma mesma forma, dependendo do contexto discursivo-pragmático e da intenção do falante” (p. 68). De acordo com Ikeda (1980, p. 122), “as diferentes formas de indeterminação nem sempre se correspondem, como querem alguns, quer no que se refere às pessoas nela incluídas, quer no tocante ao registro lingüístico”. Desta forma, os recursos só equivalem entre si, num determinado contexto de fala, se estes envolverem as mesmas pessoas do discurso.

Nos nossos dados, podemos observar que há alternância entre as formas de indeterminação, porém a equivalência só é possível quando os recursos têm o mesmo valor referencial. Comparemos os exemplos (29) e (30):

(29) Aí **você** passa na... a mandioca ali no motor, aí aquela massa já sai,... (B, 9, 144)

(30) É porque .. **nós** tinha safra de mandioca, de feijão. (P, 1, 29)

Em (29), temos um exemplo de indeterminação completa com o pronome *você*. No contexto, esta forma refere-se a qualquer pessoa que passe a mandioca no motor. Podemos substituir esta forma por qualquer outro recurso de indeterminação que o conteúdo referencial de cada um deles não é alterado, sempre vai fazer referência a qualquer pessoa. Vejamos:

(29a) Aí **a gente** passa na... a mandioca ali no motor, aí aquela massa já sai,...

(29b) Aí **nós** passa na... a mandioca ali no motor, aí aquela massa já sai,...

(29c) Aí **Ø** **passa** na... a mandioca ali no motor, aí aquela massa já sai,...

(29d) Aí **se** passa na... a mandioca ali no motor, aí aquela massa já sai,...

Ao contrário do que ocorre em (29), se substituirmos *nós* de (30) por outros recursos, há alteração na referência. No exemplo, quando o informante usa *nós*, ele está se referindo a ele mais às pessoas da comunidade, excluindo o interlocutor. Assim, na medida que substituimos um recurso por outro, haverá alteração na referência do pronome, como podemos ver abaixo:

- (30a) É porque... **você** tinha safra de mandioca, de feijão.
 (30b) É porque... **Ø tinha** safra de mandioca, de feijão.
 (30c) É porque... **Ø tinham** safra de mandioca, de feijão.
 (30d) É porque... **eles** tinha safra de mandioca, de feijão.

Vemos, então, que o conteúdo referencial de *nós* é modificado quando este é substituído por outros recursos de indeterminação. Ao alternar *nós* com *você*, $\emptyset+V3PS$, $\emptyset+V3PP$ ou *eles*, a referência deixa de ser ao informante e às pessoas da comunidade.

Em vista disso, podemos então, como afirma Cavalcante (1998), substituir um recurso de indeterminação por outro, porém havemos sempre de esperar que o efeito causado por cada um dos recursos seja diferenciado.

3.3.5 Conclusão

Feitas essas considerações, vemos que, quando se trata de indeterminação, o que se indetermina é a referência do sujeito e, para isso, ela depende essencialmente de contexto para a sua realização. Em nosso trabalho, a indeterminação é concebida como o mecanismo lingüístico que possibilita ao falante ultrapassar a determinação da referência à primeira, segunda ou terceira pessoa para um nível genérico de abrangência. Neste caso, pode envolver indistintamente as três pessoas do discurso, equivalendo a ‘todos’, ou ‘qualquer pessoa’, a depender do contexto em que é realizado.

4 METODOLOGIA

4.1 OS *CORPORA*

Os *corpora* analisados neste trabalho foram constituídos através de amostras de fala gravadas nas comunidades rurais afro-descendentes de Piabas, localizada no município de Caém, entre final de 1994 e meados de 1996, e de Bananal / Barra dos Negros, comunidades gêmeas localizadas no município de Rio de Contas, entre 1997 e 1999. São constituídos de 24 (vinte e quatro) entrevistas, sendo 12 (doze) de cada comunidade, realizadas com membros das comunidades, feitas de acordo com as técnicas da pesquisa sociolinguística (LABOV, 1983; TARALLO, 1990). Estão estratificadas da seguinte forma:

- Faixa 1²⁰ – 4 homens e 4 mulheres²¹
- Faixa 2 – 4 homens e 4 mulheres
- Faixa 3 – 4 homens e 4 mulheres

As amostras que compõem os nossos *corpora* fazem parte do Projeto “A Língua Portuguesa no Semi-Árido Baiano”, que será sumariamente apresentado abaixo.

4.1.1 O projeto “A Língua Portuguesa no Semi-Árido Baiano”

O projeto “A Língua Portuguesa no Semi-Árido Baiano” foi desenvolvido na Universidade Estadual de Feira de Santana, no Departamento de Letras e Artes, sob a coordenação das professoras Norma Lúcia Fernandes de Almeida e Zenaide de Oliveira Novais Carneiro. Tem por objetivo “contribuir para o conhecimento da realidade linguística brasileira e, de forma específica, sobre a língua falada em áreas rurais do semi-árido baiano” (ALMEIDA e CARNEIRO, 2002a, p. 11).

A idéia de criação desse projeto se deu numa apresentação em uma mesa redonda sobre o português no sertão baiano, coordenada pela professora Myrian Barbosa da Silva, na IV SBPC Especial. As primeiras gravações foram feitas em 1993 em diversas localidades do interior da Bahia, como Alto do Capim (Mucambo dos Negros), Boa Vista do Tupim, Barra e Morro do Chapéu, constituindo *corpus* piloto.

²⁰ As idades correspondentes a cada faixa etária estão descritas no item 4.4.3.2.2.

²¹ Selecionamos dois informantes de cada gênero para cada faixa etária em cada comunidade analisada, sendo doze inquéritos de cada comunidade e vinte e quatro no total.

A amostra desse projeto foi constituída entre 1994 e 2000. Atualmente conta com um banco de dados da língua falada em comunidades de diversas micro-regiões econômicas baianas, representativas da região semi-árida, a saber:

- 1) Nordeste: Casinhas, Tranqueira, Os Pereios, Lagoa do Inácio, em Jeremoabo.
- 2) Chapada Diamantina: Barra dos Negros, Bananal e Mato Grosso, em Rio de Contas.
- 3) Paraguassu: Matinha, São José das Itapororocas, em Feira de Santana.
- 4) Piemonte da Diamantina: Piabas, em Caém, atualmente Ancelino da Fonseca.
- 5) Sudoeste: Cinzento, em Vitória da Conquista.

O projeto “A Língua Portuguesa no Semi-Árido Baiano” faz parte do acervo do Núcleo de Estudos em Língua Portuguesa Rural (NELPRU), que só foi criado enquanto núcleo em 1998. O NELPRU teve dois objetivos gerais: a formação de dados de registros fônicos e a análise lingüística da fala das comunidades rurais do interior da Bahia.

Pensou-se em alargar a região de alcance do projeto, a partir da idéia de que esse núcleo pudesse congregiar projetos individuais que surgissem a partir da coleta de dados em diferentes regiões. Assim, o NELPRU foi dividido em dois núcleos: 1) *A Língua Portuguesa no Semi-árido Baiano* e 2) *A Língua Portuguesa no Recôncavo Baiano*. Esses *corpora* foram constituídos com o objetivo de dispor de amostras que possam ser usadas em estudos lingüísticos que tenham finalidades diversas, como, por exemplo, em dissertações de mestrado e teses de doutorado.

4.1.2 As comunidades

Como foi dito, a amostra que compõe os nossos *corpora* faz parte do projeto “A Língua Portuguesa no Semi-Árido Baiano”. Das localidades que compõem este projeto, selecionamos as comunidades de Piabas e Bananal / Barra dos Negros²² para análise do fenômeno da indeterminação do sujeito.

A localidade de Piabas é um pequeno povoado, de maioria negra, que faz parte da zona rural de Ancelino da Fonseca ou Caém, como é mais conhecido, antigo distrito de Jacobina.

²² Faremos referências a essas comunidades como se fossem apenas uma, pois as entrevistas que constituem a amostra de Rio de Contas foram feitas nas duas comunidades, sem distinção de uma ou de outra. Além do mais, ambas distanciam-se por apenas 2 km, tendo, portanto, características histórico-sociais e lingüísticas em comum.

Caém foi fundada no final do século XIX e está localizada na encosta da Chapada Diamantina, a 329 km de Salvador. Conta com uma população de 12.563²³ habitantes.

A cidade foi criada quando da corrida às minas de ouro de Jacobina foram construídas algumas residências e uma igreja, aproveitando o vale fértil e seguro, às margens do Rio da Prata, hoje denominado Rio Caém, antes da subida das serras, ganhando a denominação de Sítio Papagaio.

Com o início da construção da estrada de ferro, ligando Senhor do Bonfim a Jacobina, muitos dos trabalhadores que participaram da empreitada acabaram fixando residência no Sítio Papagaio e aproveitando a fertilidade do vale. Um destes trabalhadores foi o Engenheiro francês Henry D'Brutelles que era responsável pela construção da estrada de ferro e que denominou a povoação com o nome atual de Caém, uma homenagem à sua esposa Madame Marta D'Brutelles, natural de Caen na França.

Bananal e Barra dos Negros são comunidades rurais afro-descendentes separadas por apenas dois quilômetros. Ambas foram fundadas no século XVII por escravos africanos que sobreviveram de um naufrágio na costa baiana, no qual os sobreviventes nadaram até a praia e, tendo o curso do Rio das Contas como guia, adentraram sertão acima em busca de um lugar seguro para se estabelecerem. Escolheram as cabeceiras do Rio Brumado e por lá ficaram, cultivando suas roças, mantendo suas culturas, suas tradições. Nelas vivem apenas negros descendentes de escravos que se agruparam por essas áreas nos séculos XVII e XVIII.

O povoado de Bananal fica no vale do Rio Brumado ou Vale do Brumadinho, como é chamado pelas pessoas da região. Atualmente, conta com cerca de 38 casas distribuídas de forma esparsa em condições de urbanização precárias. Pratica lavoura de subsistência e tem um relativo isolamento. Apesar das raízes negras, não se têm relatos de cultos de candomblé, lá predomina a religião católica e a igreja do povoado é a de São José.

O povoado de Barra é um pouco maior do que o de Bananal, sendo também de composição negra. Possui uma estrutura maior, com maior número de residências. Predomina também a religião católica e a igreja é a de São Sebastião, a qual foi construída no século XIX.

De acordo com a Fundação Palmares, as comunidades, juntas, possuem por volta de 740 habitantes. Deste total, apenas metade disso é de residentes fixos, pois há muitas migrações sazonais, as pessoas vão para São Paulo, depois voltam para o Nordeste, descem de novo para o sul e retomam ao norte num vai e vem sem fim.

²³ Censo 2000.

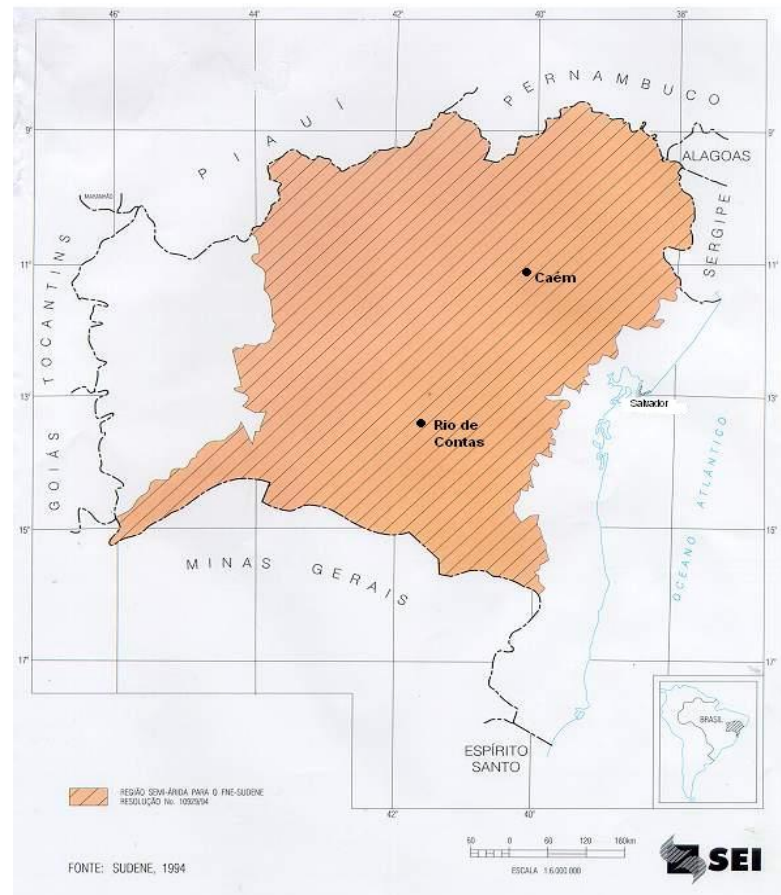
Ambas as comunidades fazem parte do município de Rio de Contas. Rio de Contas fica no sudoeste da região do Parque Nacional da Chapada Diamantina, a 730 quilômetros da capital baiana, e tem uma população de 13.935 habitantes²⁴. Seus primeiros moradores foram escravos alforriados que se reuniram no Arraial dos Creoulos. No início do século XVIII, com a chegada de bandeirantes interessados em novas regiões de exploração do ouro, inicia-se um ciclo que marcou a história da região. Nessa época, um novo arraial (hoje chamado de Mato Grosso) foi fundado, atraindo mais pessoas. Também nessa época chegaram os padres jesuítas.

Por causa do ouro, o arraial cresceu rapidamente e passou a vila em 9 de fevereiro de 1724, passando a chamar-se Vila Nova de Nossa Senhora do Livramento das Minas do Rio de Contas. Mais tarde, em 1742, por ordem do conde Galvéas, a sede da vila foi transferida para um ponto situado duas léguas acima, local onde hoje está a cidade de Rio de Contas, antigo Pouso dos Creoulos.

Na figura 1, abaixo, podemos ver o mapa da região semi-árida baiana e a localização dos municípios dos quais as comunidades rurais que estamos analisando fazem parte.

²⁴ CENSO 2000.

Figura 1: Mapa da Região Semi-árida Baiana



Fonte: Sudene, 1994.

4.1.3 Os informantes

Os informantes escolhidos para as entrevistas são moradores das localidades selecionadas, filhos de pais que nasceram na mesma localidade, que não se afastaram por mais de 10 anos consecutivos da cidade de origem. Eles têm no máximo quatro anos de escolarização, sendo a maioria deles totalmente analfabeta.

4.1.4 As entrevistas

As entrevistas que compõem os nossos *corpora* foram realizadas com membros das comunidades, feitas de acordo com as técnicas da pesquisa sociolinguística (LABOV, 1983; TARALLO, 1990). São do tipo DID (diálogo entre informante e documentador) e possuem em média 50 minutos de gravação cada uma. Elas caracterizam-se por uma conversação espontânea que se instaura entre o documentador e o informante já que o intuito da recolha foi

de gravar a fala vernácula da comunidade. As entrevistas foram realizadas na casa dos entrevistados e no comércio local, fazendo com que estes se sentissem livres para falar à vontade. Não foi possível evitar a presença de circunstâncias, porém estes foram admitidos, pois, assim, criava-se uma situação de fala mais coloquial.

Os inquéritos, gravados em situações informais, como dissemos, apresentam uma variedade de assuntos que, em linhas gerais, versam sobre a história da fundação da(s) localidade(s), a seca na região, a situação econômica, política e social do lugar, a família, histórias antigas, religião, assuntos domésticos, dentre outros, constituindo, assim, um conhecimento da vida e da cultura desses informantes.

4.1.5 A transcrição

A transcrição das entrevistas foi feita ortograficamente, o que não exige os pesquisadores de recorrerem às fitas, principalmente e obrigatoriamente, quando forem fazer trabalhos no âmbito da fonética e da fonologia. Esta transcrição foi realizada a partir de uma norma elaborada pelas coordenadoras do projeto, sendo que a elaboração destas normas teve como base duas outras: a do Projeto Vestígios de Dialectos Crioulos em Comunidades Afro-Brasileiras Isoladas²⁵ e a do Projeto NURC.

4.2 SUPORTE QUANTITATIVO

No estudo da variação, teremos que ter em mente que o seu pressuposto básico no uso da língua é o de que a heterogeneidade não é aleatória, assim como a homogeneidade também não é, mas regulada por um conjunto de regras que funcionam para favorecer ou desfavorecer o uso de uma ou de outra forma em cada contexto.

Cabe a nós pesquisadores delimitarmos o fenômeno lingüístico variável, a variável dependente, e identificar suas variantes. Fazendo isso, é fundamental identificarmos, também, as circunstâncias que tendem a influenciar o uso de uma ou de outra variante. Teremos então as variáveis independentes, que podem ser internas ao sistema lingüístico ou externas a ele²⁶.

²⁵ O *Projeto Vestígios de Dialectos Crioulos em Comunidade Rurais Afro-Brasileiras Isoladas* foi implantado no Departamento de Letras Vernáculas do Instituto de Letras da UFBA, em fevereiro de 1993, e contou com o financiamento do Australian Research Council e com bolsistas do Iniciação Científica do Convênio UFBA/CNPq. Atualmente conta com a coordenação do Prof. Dr. Dante Lucchesi, em associação com o Prof. Dr. Alan Baxter, da Universidade de La Trobe, em Sidney - Austrália.

²⁶ Cf. o item 2.1 deste trabalho.

Segundo Naro (2003, p. 16-7),

o problema central que se coloca para a Teoria da Variação é a avaliação do *quantum* com que cada categoria postulada contribui para a realização de uma ou de outra variante das formas em competição. No uso real da língua, que constitui o dado do lingüista, seja na forma falada ou na forma escrita, tais categorias se apresentam sempre conjugadas; na prática, a operação de uma regra variável é sempre o efeito da atuação de vários fatores.

Tendo como base esse pressuposto da Teoria da Variação colocado por Naro, foram desenvolvidos modelos matemáticos²⁷ que possibilitam o tratamento estatístico dos dados lingüísticos. São eles: o modelo aditivo (LABOV, 1969), o modelo multiplicativo (CEDERGREN; SANKOFF, 1974) e o modelo logístico ou misto (ROUSSEAU; SANKOFF, 1978). Atualmente, o mais usado na análise dos fenômenos variáveis é o logístico ou misto, que apresenta uma síntese dos modelos anteriores.

Para a obtenção dos resultados estatísticos que subsidiaram esta pesquisa, utilizamos o pacote de programas VARBRUL – do inglês *variable rules* –, criado por David Sankoff (cf. SANKOFF, 1988; PINTZUK, 1988)²⁸, que calcula os pesos relativos dos fatores em relação à variável dependente, a partir das frequências brutas, e seleciona os grupos de fatores mais favoráveis a aplicação da regra.

O pacote VARBRUL é constituído por um conjunto de programas que podem ser divididos em três grupos:

- 1) os que preparam os dados para análise – *checktok*, *readtok* e *makcell* ou *make3000*;
- 2) os que fornecem os pesos relativos – *ivarb* ou *varb2000* (análise binária), *tvarb* (análise ternária) e *mvarb* (análise eneária);
- 3) e os que efetuam tarefas de apoio – *tsort*, *texsort* e *crosstab* ou *cros3000*.

A análise quantitativa realizada por esse conjunto de programas só é feita a partir da codificação das ocorrências da variável dependente analisada com base nos valores atribuídos às variáveis independentes que são selecionadas previamente pelo pesquisador.

O primeiro resultado fornecido pelo VARBRUL são as frequências de uso de cada variante do fenômeno em análise (variável dependente) associada a cada um dos fatores das variáveis independentes. A partir daí, a análise quantitativa prossegue trabalhando com diversos níveis de análise, num processo gradual e crescente (denominado *step up*), entre as variáveis independentes e atribuindo pesos relativos para os seus respectivos fatores. Cada

²⁷ Em Naro (2003) há uma clara exposição do desenvolvimento dos modelos matemáticos.

²⁸ Encontramos em Scherre (1992/1993) e Brescancini (2002) uma análise mais detalhada do funcionamento do VARBRUL.

nível é definido a partir da escolha de uma variável dependente cujos resultados serão combinados com os resultados obtidos nas demais variáveis até que não reste nenhuma variável estatisticamente relevante.

Para a seleção das variáveis estatisticamente relevantes, o programa trabalha com um nível de significância e com um teste de verossimilhança máxima (denominado *log likelihood*). Esta seleção significa que a variação existente nos dados não é aleatória, e sim que, do ponto de vista estatístico, a variável independente dá conta de parte da variação que está sendo estudada.

Após o término da seleção dos grupos estatisticamente significativos, o programa passa a trabalhar de forma inversa (método *step down*), relacionando todos os fatores da variável independente com a variável dependente e eliminando os grupos que desfavorecem a aplicação da regra. Por fim, o programa indica os grupos não relevantes para o fenômeno.

Os programas que compõem o VARBRUL geram como produto final o peso relativo. É através dele que as variáveis são selecionadas como estatisticamente relevantes. Os pesos relativos medem, numa escala de zero a 0,99, a influência de cada fator das variáveis independentes sobre a variável analisada. Se estes pesos se apresentam acima de 0,50, são interpretados como favorecedores a aplicação da regra (a realização da variante em foco); abaixo de 0,50, como desfavorecedores. Todos os valores próximos de 0,50 são considerados neutros.

Os resultados finais também devem estar dentro de uma margem definida pelo nível de significância que vai lhes conferir confiabilidade estatística. O nível de significância usado nos programas do VARBRUL é de 0,050, o que significa que o nível de segurança dos resultados finais, e também dos resultados de cada variável dependente no momento da sua seleção, deve ser igual ou inferior a 0,050.

É importante ressaltar que os resultados numéricos obtidos pelos programas só têm valor estatístico, sendo um instrumento valioso que pode nos auxiliar a entender um pouco mais o comportamento de fenômenos lingüísticos. Cabe ao lingüista interpretar e atribuir o seu valor lingüístico.

4.3 LEVANTAMENTO DAS OCORRÊNCIAS

O levantamento das ocorrências do fenômeno estudado não foi feito por nós diretamente nos textos orais, isto é, através da audição das fitas, mas no texto já transcrito das entrevistas, visto que as amostras de fala das duas comunidades que estamos analisando já

foram editadas em forma de livro. O conjunto de ocorrências que constituem a base de dados da presente pesquisa foi depreendido dos textos impressos.

4.3.1 Critérios de seleção das ocorrências

Na seleção das ocorrências de sujeito indeterminado estabelecemos algumas restrições. Assim, não consideramos:

- a) Os casos em que o informante usou dois recursos de indeterminação seguidos para a mesma ocorrência.

(31) Inf.: (...) **Nós hoje, a gente** sai tudo ali do banhe. Oito hora da noite, nove hora, em casa, a gente tá dento d'água tomano banhe. Se num tomar, num consegue dormir de noite. (B, 12, 212)

- b) As ocorrências em que a forma indeterminada estava inserida no discurso de terceiros.

(32) Inf.: Mas cadê, num vem. Eles diz: '**mas a gente vem aqui, vamo fazer isso aí**', mas só que num faz (P, 5, 103);

- c) As formas repetidas em estruturas idênticas (consideramos apenas a primeira ocorrência).

(33) Doc.: Eu ouvi dizer que era antigamente, né, quando o rio enchia ninguém passava pra cá?

Inf.: **Passa não**, quano o rio enche aí, **não passa não**.

Doc.: Ainda não passa?

Inf.: **Passa não**.

Doc.: Quando esse rio enche, né?

Inf.: É quano esse aí enche, não passa não, **não passa ninguém**. (B, 4, 55)

- d) As estruturas esboçadas e abandonadas com posterior reformulação.

(34) Inf.: É vazio. Tem hora que **a gente fica... Já acostudou**, é um lugar muito sossegado, graças a Deus. (B, 9, 135)

- e) Os dados em que a transcrição revela que houve interrupção, truncamento ou indecisão.

(35) Doc.: E a farinha, como é que faz?

Inf.: A farinha também é aquele mermo tipo da... da rapadura, né? **Você tem... quem tem o motor**, hoje em dia é mai fáci, né? (B, 9, 143-4)

- f) Os verbos impessoais, pois não fazem referência a um sujeito;

(36) **Chove** muito lá fora.

- g) As repetições pelo informante da forma indeterminada contida na estrutura empregada pelo documentador antes da resposta do informante.

(37) Doc.: Num perdia. Perdeu tudo foi?

Inf.: Foi. **Perdeu tudo**. (B, 3, 20)

4.4 VARIÁVEIS CONSIDERADAS

Descreveremos neste item as variáveis dependentes e independentes estabelecidas para esta pesquisa bem como apresentaremos as hipóteses postuladas para o estudo da atuação dos grupos de fatores no condicionamento do uso das formas de indeterminação do sujeito. Essas hipóteses se fundamentam na observação dos *corpora* e na literatura lingüística sobre o tema em estudo.

4.4.1 Variável dependente

A indeterminação do sujeito em português não é expressa por apenas dois recursos como é colocado na GT. Os estudos mostram que diversos recursos são utilizados pelo falante para indeterminar o sujeito. Além das formas canônicas, $\emptyset+V3PP$ e $\emptyset+V+SE$, consideramos outras formas utilizadas pelo falante como recurso indeterminador, dentre elas: *nós*, *a gente*, *você*, *eles*, $\emptyset+V3PS$, $\emptyset+VINFL$. Passemos agora a descrever em que consiste cada uma delas e os critérios de seleção destas variantes nos *corpora*.

4.4.1.1 Nós

A forma *nós* é classificada, segundo a GT, como um pronome pessoal sujeito usado na primeira pessoa do plural, ou seja, como a primeira pessoa do discurso (quem fala). A partir da análise das GTs, verificamos que os autores relacionam outros usos a esta forma, a exemplo de Cunha e Cintra (2001) que no item referente à ‘extensão de emprego dos pronomes retos’ dizem que “na linguagem formal, certos pronomes retos adquirem valores especiais” (p. 283-284). Para abordar esses valores, os autores citam o “plural de modéstia”, no qual o pronome *nós* é usado por escritores e oradores no lugar da primeira pessoa do singular, *eu*, evitando o tom de imposição ou de pessoalidade de suas opiniões, e o “plural de majestade”, usado pelos reis de Portugal como símbolo de grandeza e poder de suas funções e que é usado até hoje pelos altos dignitários da igreja.

De modo geral, as GTs apresentam o pronome *nós* como forma plural de ‘eu’ (eu + eu), deixando de lado o seu uso mais abrangente de um ‘eu ampliado’. Segundo Benveniste (1995), quando o falante se coloca no discurso juntamente com uma ou mais pessoas, o pronome se apresenta com diferentes conteúdos. Assim, o pronome *nós* pode abranger, além do ‘eu’ e do ‘não-eu’, a ‘não-pessoa’ (*alia*), compreendendo, assim, três possibilidades de ampliação do ‘eu’: ‘eu’ + ‘não-eu’, ‘eu’ + ‘não-pessoa’ e ‘eu’ + ‘não-eu’ + ‘não-pessoa’ (ALBÁN; FREITAS, 1991a).

Na análise que fizemos nas GTs, a única que vimos a concepção de ‘ampliação do eu’, conforme coloca Benveniste (1995), foi a de Bechara (1999), na qual o autor, ao abordar os pronomes pessoais, especificamente o pronome *nós*, diz que “o plural nós indica ‘eu’ mais outra ou outras pessoas, e não ‘eu + eu’” (p. 164).

Desta forma, observamos que essa ‘ampliação do eu’ comporta diferentes graus de generalização²⁹, variando numa escala que vai do grau mais determinado para o menos determinado, incluindo sempre o comprometimento do ‘eu’. Isso nos permite verificar que esta é também mais uma possibilidade usada pelo falante para indeterminar o sujeito dos seus enunciados, concepção que não é abordada em nenhuma das gramáticas analisadas.

Assim, o pronome *nós*, além de ser utilizado pelo falante para referir-se a si próprio ou a um grupo do qual faz parte, pode representar também um grupo indeterminado de pessoas cuja referência não se pode ou não se deseja recuperar. Nos trabalhos sobre indeterminação, observamos o uso de *nós* também como uma das possibilidades de se indeterminar o sujeito

²⁹ Já falamos sobre grau de indeterminação no item 3.3.3.

(MILANEZ, 1982; ROLLEMBERG, 1991; CUNHA, 1993; SETTI, 1997; GODOY, 1999; dentre outros).

Nos nossos dados, o mesmo foi verificado, encontramos *nós* como recurso para indeterminar o sujeito, conforme podemos ver nos exemplos (38) a (40):

(38) Doc.: O tempo tá passando mais rápido?

Inf.: Tá, pois se **nós** tamo no fim da era. É de pegar, vai passano mermo aí, porque é pá nós ir embora tudo. (P, 12, 172)

(39) Inf.: Avoador é aqueles palitinho branco assim que... uns palitinho branco que nós... que **nós** come aqui, avoador, né? (B, 3, 27)

(40) Doc.: Quer dizer que o senhor gosta de ir pra romaria? Pra Lapa?

Inf.: Todo ano eu vou. (...) **Nós** levava rapadura, carne. Premêro **nós** levava galinha, matava duas galinha, fritava e agora botava na mochila, e agora viajava (P, 13, 167)

O emprego desta forma pelos falantes das comunidades analisadas varia de um grau mais indeterminado até um grau menos indeterminado. Eles usam *nós* para referir-se a qualquer pessoa, como no exemplo (38), às pessoas que vivem na comunidade de Bananal, como no exemplo (39), e também para referir-se às pessoas que foram para a romaria em Bom Jesus da Lapa, como é possível identificar na fala do informante no exemplo (40).

4.4.1.2 A gente

Diversos estudos vêm mostrando que a forma *a gente* passou a fazer parte do sistema pronominal do português como indicador da primeira pessoa do plural, variando com o pronome *nós*.

Originariamente, o substantivo “gente” advém do substantivo feminino latino *gēns*, *gēntis* que significa ‘raça’, ‘família’, ‘tribo’, ‘o povo de um país, comarca ou cidade’. Ao lado do artigo definido feminino singular “a”, resultou na forma *a gente* que, atualmente, tem o seu emprego ampliado, equivalendo à função de pronome de primeira pessoa do plural, em variação com o pronome *nós*, com o verbo normalmente com flexão similar à de terceira pessoa do singular.

Embora a forma *a gente* seja reconhecida como referência à primeira pessoa do discurso, em geral, os gramáticos não levam em consideração esta variante. Nas gramáticas

que pesquisamos, é unânime tê-la com função de pronome, porém sua classificação é diversificada entre os autores. Encontramos as seguintes classificações:

a. Pronome de Tratamento

Almeida (1969, p. 172), ao conceituar pronomes de tratamento, diz:

Chama-se pronomes de tratamento as palavras e expressões que substituem a terceira pessoa gramatical: fulano, beltrano, sicrano, **a gente** [grifo nosso], você, vossa mercê, vossa senhoria, sua santidade, sua majestade.

Bechara (1992, p. 166) faz a seguinte observação no final do capítulo sobre pronomes:

O substantivo **gente**, precedido do artigo **a** e em referência a um grupo de pessoas em que se inclui a que fala, ou a esta sozinha, passa a pronome e se emprega fora da linguagem coloquial cerimoniosa. Em ambos os casos o verbo fica na 3ª pessoa do singular.

b. Locução Pronominal Indefinida

Kury (1989), ao abordar os pronomes indefinidos, faz referência à forma *a gente* como uma locução:

Há pronomes indefinidos constituídos de mais de uma palavra, - são as LOCUÇÕES PRONOMINAIS INDEFINIDAS: o mais, cada um, cada qual, quem quer, **a gente** [grifo nosso] etc. (p. 151)

Melo (1980, p. 122), quando menciona as modalidades de sujeito indeterminado, também apresenta essa variante como “expressão de valor indefinido”.

c. Fórmula de Representação da 1ª Pessoa

Cunha e Cintra (2001, p. 296) registram:

No colóquio normal, emprega-se **a gente** por **nós** e, também, por **eu**:
 Houve um momento entre nós
 Em que **a gente** não falou. (F. Pessoa).

Você não calcula o que **a gente** ser perseguida pelos homens.
 Todos me olham como se quisessem devorar-me. (C. dos Anjos)

Em nenhuma das gramáticas analisadas encontramos a forma *a gente* como pronome pessoal nem como recurso usado para indeterminar o sujeito. Porém, o ingresso da expressão *a gente* no quadro dos pronomes pessoais já motivou algumas pesquisas tanto na fala culta quanto na popular, tanto para mostrar que esta é uma forma variante de *nós* (OMENA, 1986; FREITAS, 1991; LOPES, 1993; MACHADO, 1995) quanto como recurso utilizado pelo falante para indeterminar o sujeito (CUNHA, 1993; SETTI, 1997; GODOY, 1999).

Em nossos dados, a forma *a gente* é a mais utilizada pelos falantes para indeterminar o sujeito. E, assim como a forma *nós*, comporta diferentes graus de generalização, podendo envolver o ‘não-eu’ e a ‘não-pessoa’ determinadas ou indeterminadas pelo contexto, a depender da ‘ampliação do eu’. Alguns exemplos retirados dos nossos dados mostram o uso dessa forma pelos informantes:

(41) Inf.: Pra tirar água. Aí agora quando enxuga aí **a gente** cessa na peneira, na peneira. Aí agora, **a gente** pisa a massa no pilão. E aí agora, é só fazer beiju. (P, 10, 149)

(42) Inf.: No mês de mauço, que é o mês que a gen... **a gente** chama aqui mês da quaresma, né? (B, 3, 20)

(43) Doc.: Mas no seu tempo era outro, né, brincava?

Inf.: **A gente** brincava muito de boneca com as amiga, com as vizinha. Até hoje eu ainda tenho saudade de minhas boneca. [rindo](B, 4, 49,)

4.4.1.3 Você

É freqüente o emprego do pronome sujeito de segunda pessoa do singular *tu* entre os falantes do português do Brasil. A maioria dos estudiosos que analisa a origem do pronome *você* na língua portuguesa (CINTRA, 1972; FARACO, 1996; TEIXEIRA, 2001) afirma que ele é uma “simplificação” do antigo pronome de tratamento *Vossa Mercê*, que começou a ser utilizado no século XIV em Portugal como forma de tratamento própria para o rei. Esta forma, no decorrer do tempo, passa a ser usada como forma de tratamento para várias pessoas e não somente mais para o rei, passando por um processo de vulgarização, sendo assimilada rapidamente pelo povo. Como diz Said Ali (1966, p. 93), “nasceu em bôca do povo a variante *você*”.

A transformação de *Vossa Mercê* passa por etapas intermediárias, surgindo diversas variantes. De acordo com Coutinho (1958, p. 280), a transformação foi: **Vossa Mercê** >

Vossemecê > Vosmecê > Você. Assim, vemos que o pronome *você* nasce da transformação fonética e semântica de *Vossa Mercê* que é fortemente influenciada pelas relações sociais estabelecidas em determinadas épocas, principalmente nos séculos XV e XVI.

De modo geral, as GTs classificam a forma *você* como pronome de tratamento de segunda pessoa, utilizado como forma íntima e familiar de tratamento, que leva o verbo à terceira pessoa. Alguns autores a definem como ‘forma de tratamento indireto’, o que é uma maneira de explicar a irregularidade de um pronome de segunda pessoa concordar com o verbo na terceira (LUFT, 1996, p. 117; BECHARA, 1999, p. 165). Nas gramáticas atuais³⁰, essa forma é classificada, no geral, como pronome pessoal ou como pronome de tratamento, não sendo definida claramente a sua função. Nestas gramáticas, os autores também abordam o emprego desta forma como substituta de *tu*.

Também são pronomes de tratamento **o senhor, a senhora e você, vocês. O senhor e a senhora** são empregados no tratamento cerimonioso; **você, vocês**, no tratamento familiar. **Você, vocês** são largamente empregados no português do Brasil, praticamente substituindo as formas **tu** e **vós**. (INFANTE, 1995, p. 89)

Em nenhuma das GTs analisadas o uso da forma *você* é considerado “como expressão de um conteúdo mais genérico, de um conjunto constituído das duas pessoas do discurso mais a ‘não-pessoa’” (ROLLEMBERG *et al*, 1991, p. 41), o que configura como um dos recursos utilizados pelo falante para indeterminar o sujeito.

Em nossos dados, o pronome *você* aparece não só para referir-se à segunda pessoa do discurso como também à primeira pessoa mais a segunda pessoa mais ‘todo mundo’, indistintamente, tendo, assim, uma característica indeterminadora. Nesse caso, *você* é indeterminado quando o falante se reporta a um conteúdo genérico.

4.4.1.4 Eles

De acordo com a GT, *eles* é um pronome pessoal da terceira pessoa do discurso (de quem se fala), sendo expresso em sua forma no singular, ‘ele’. Além dessa concepção, encontramos algumas informações complementares em relação ao seu uso, porém nenhuma delas cita o pronome *eles* como uma das possibilidades de o falante indeterminar o sujeito.

³⁰ Consideramos como gramáticas atuais àquelas tiveram sua primeira edição publicada a partir da década de 90.

Embora restrita à terceira pessoa, ou à ‘não-pessoa’, que segundo Benveniste (1995, p. 254) “designa o conjunto de seres não-pessoais”, a forma *eles* é estreitamente abrangente, representando qualquer pessoa, por oposição às duas primeiras pessoas, que se referem especificamente aos interlocutores.

A forma pronominal *eles* é usada para se indeterminar o sujeito em português quando não há referências anteriores no contexto com as quais se possa relacioná-la. Segundo Menon (1994), o uso de *eles* na indeterminação é a forma plena da tradicional $\emptyset + V3PP$. Para ela, não há distinção no uso de uma ou de outra forma e sua presença pode ser uma consequência do fato de a língua portuguesa está em vias de preencher a posição de sujeito nulo. Discordamos desse posicionamento, pois, como mostraremos na análise, tais formas têm valores diferentes na hora de indeterminar o sujeito.

Assim como Milanez (1982) e Godoy (1999), observamos que uma das características apresentadas à forma de indeterminação *eles* é o seu envolvimento com um grupo delimitado socialmente, implícito ou explícito no contexto, sendo esta forma usada para distinguir o grupo que o falante pertence do grupo que ele se refere, como mostram os exemplos (44) e (45):

(44) Doc.: Dentro da cidade também, lá pa Feira de Santana, pa Salvador, eles chegam prometendo mundos e fundo.

Inf.: Mundos e fundo, sem ter condições de dar, ta entendeno?

Doc.: Né?

Inf.: Quer dizer que condições **eles** têm, né?

Doc.: Claro que têm.

Inf.: Mais **eles** promete... promete... promete e aí \emptyset vai enrolano... enrolano, é. (P, 3, 71)

(45) Doc.: E assim pra médico, como é?

Inf.: Rio de Conta.

Doc.: É?

Inf.: É.

Doc.: Resolve tudo por lá mermo?

Inf.: É. Resorve por lá mermo. Quando **eles** num pode dar jeito, \emptyset manda pra Livramento, né? (B, 35, 285)

No exemplo (44), quando documentador e informante conversam sobre política, percebemos a oposição entre políticos e eleitor. No exemplo (45), há oposição entre médico e paciente.

O que verificamos nos nossos *corpora*, em relação ao uso da forma *eles* para indeterminar o sujeito, não é apenas com esse efeito de distinção de grupos sociais, encontramos, ainda, exemplos com esta forma de indeterminação em que não verificamos essa oposição ao grupo do informante, como no exemplo (46):

- (46) Doc.: Pois lá em Feira, a gente compra as coisas assim, fazem até mal.
 Inf.: Faz mermo, porque **eles** colocam muito produto. (B, 4, 59)

Ao selecionar as ocorrências com esta forma indeterminadora, estabelecemos o seguinte critério: se a primeira aparição foi de *eles*, os zeros correspondentes foram considerados anáfora desse *eles*; se a primeira aparição foi de \emptyset +V3PP ou \emptyset +V3PS, assim também foram consideradas as formas seguintes, no mesmo período.

4.4.1.5 \emptyset +V3PS

Nas gramáticas analisadas, apenas Bechara (1992) faz referência a esta forma como uma das possibilidades de indeterminarmos o sujeito em português. A forma verbal de 3ª pessoa do singular sem sujeito expresso também pode representar o sujeito indeterminado, desde que não seja possível recuperar o seu referente no contexto com uma forma plena. Assim, embora provida de uma forma de 3ª pessoa desinencial, seu conteúdo abrange as três pessoas do discurso, daí o seu caráter indeterminador.

Em nossos dados, esta foi a segunda forma mais recorrente. Podemos ver um exemplo em (47):

- (47) Doc.: Como é que a senhora faz doce de mandioca?
 Inf.: \emptyset **Rela** mandioca, \emptyset **lava** muito bem lavada, \emptyset **tira** a goma dela todinha, \emptyset **passa** na peneira, \emptyset **côa**, \emptyset **põe** pa freventar, depoi de freventada \emptyset **torna** levar na água, \emptyset **lava** te tira tudo quanto é escorrego. Aí, fica meia crespa. Aí, cê faz a calda lá na panela e coloca lá dento... (B, 34, 265)

Neste exemplo, quando a informante explica como se faz doce de mandioca, o sujeito pode ser qualquer pessoa, não sendo possível recuperar a referência dentro do contexto. É isso que vai diferenciar esta forma das formas que apresentam as referências expressas no contexto.

No que tange às expressões impessoais, conforme dissemos no item 4.4, estas não foram consideradas, pois, apesar de possuírem estrutura idêntica à da forma em análise

(\emptyset +V3PS), não são formas indeterminadoras, pois não possuem referência alguma no contexto.

(48) **Choveu** o dia inteiro.

Nesta oração não é possível de forma alguma identificar um referente, a idéia expressa pelo verbo não pode aplicar-se a algo ou a alguém.

4.4.1.6 \emptyset +V3PP

O verbo na terceira pessoa do plural é uma das formas consagradas pela GT para indeterminar o sujeito. Para Milanez (1982, p. 69), “à primeira vista, pode parecer que o vazio lexical desta ocorrência seja uma ausência opcional do pronome indeterminador *eles*”. A mesma autora explica que essa hipótese não é verdadeira, pois os dois recursos podem ocorrer em ambientes diferentes. Ela reforça a idéia de que *eles* só é usado para indeterminar o sujeito nos contextos em que há uma referência a um grupo social a que o sujeito pertence, já a forma \emptyset +V3PP não está sujeita a esta restrição.

Almeida (1992), ao falar do uso dos indeterminadores, diz que, ao contrário da forma *eles* que é usada para focalizar o sujeito, a forma \emptyset +V3PP “ocorre em enunciados em que o elemento mais focal é a ação/evento/processo significado pelo verbo” (p. 173). Godoy (1999) partilha do mesmo ponto de vista, dizendo que o que se evidencia no uso desta forma é o conteúdo expresso pelo verbo.

Concordamos com a análise das referidas autoras e comprovamos com exemplos retirados dos nossos *corpora*. Vejamos:

(49) Inf.: (...) Tá c’uns dois ano que \emptyset **colocar** a lâmpada na cidade que veio a energia pra gente. (B, 8, 114)

(50) Doc.: Mas essa igreja tem pouco tempo de construída, né?

Inf.: É essa igreja é nova, mair sempre teve igreja...

Doc.: Sempre teve.

Inf.: Que \emptyset **derrubar** a outra, né? \emptyset **Fizero** aquela nova. Mas teve sempre teve igreja, derde eu pequeno. (P, 7, 132-3)

Nesses exemplos, podemos observar que o foco maior é o conteúdo expresso pelo verbo, não sendo possível identificar quem realizou a ação, remetendo à generalização.

Assim, podemos perceber que a análise desta forma, assim como da forma *eles*, requer um cuidado maior. Para tanto, selecionamos o indeterminador $\emptyset+V3PP$ quando o vazio não fez referência ao pronome *eles* e quando não foi anáfora nem catáfora de um sujeito determinado.

4.4.1.7 $\emptyset+VIN$

O uso da forma $\emptyset+VIN$ para indeterminar o sujeito não é comum na tradição gramatical, apenas Luft (1996) e Cegalla (1995) a consideram. Da mesma forma que o $\emptyset+V3PS$, a ausência de um termo representando o sujeito junto a verbos no infinitivo, seja um sujeito explícito na oração anterior ou imediatamente posterior, implica indeterminação, porque fica impossível a identificação do agente verbal, remetendo a um ser genérico, indeterminado.

O responsável pelo efeito indeterminador desta forma verbal é justamente a ausência de sujeito lexical. Em (51), temos um exemplo desta forma representando a indeterminação:

- (51) Doc.: Aqui dá de tudo?
 Inf.: Tudo que \emptyset **prantar** aqui dá. (B, 4, 59)

Nesse exemplo, vemos que o infinitivo exprime a ação de maneira geral e indeterminada, sem a indicação de um ser ou de um agente a que ele possa ser referido.

Menon (1994), Setti (1997) e Godoy (1999) consideram esta forma juntamente com a forma de $\emptyset+V3PS$, do que discordamos, visto que ambos os recursos têm valores diferentes, como veremos na análise dos dados.

4.4.1.8 $\emptyset+V+SE$

Desde suas origens, o emprego básico do *se* tem sido o de pronome reflexivo (cf NARO, 1976). Com o uso, o *se* passou a diversificar-se, indicando primeiramente passividade e posteriormente assumindo posição de sujeito, indicando índice de indeterminação³¹.

³¹ De acordo com Naro (1976) e Nunes (1991), o uso do *se* indeterminador no português se deu devido a uma reanálise sintática. Com o desaparecimento do agente da passiva em construções com *se*, o agente passou a ser interpretado como um sujeito indeterminado. Além dessa mudança no estatuto sintático do *se* (de apassivador para indeterminador), Nunes (1991) aponta uma tendência no uso do *se* com infinitivo. Este uso é apontado por Galves (1987, *apud* CAVALCANTE, 2002) como uma característica do português brasileiro.

Na GT, o *se* é considerado como um pronome pessoal oblíquo, partícula apassivadora, quando acompanhado de verbos transitivos diretos, e índice de indeterminação do sujeito, quando acompanhado de verbos intransitivos, transitivos diretos e de ligação, a depender do contexto. Como índice de indeterminação, apresenta-se como um dos mais fortes indeterminadores, sendo considerado, juntamente com o verbo na terceira pessoa do singular, como uma das formas clássicas de indeterminação do sujeito.

Em nosso trabalho, consideramos como indeterminado tanto o *se* apassivador como o *se* índice de indeterminação do sujeito, ou seja, lhe atribuímos apenas uma função, ao contrário da GT que lhe atribui duas.

Em nossos dados, esta foi a forma de indeterminação menos produtiva, sendo encontradas apenas 15 ocorrências em todos os *corpora*. Eis alguns exemplos da construção com *se* retirados dos nossos dados:

(52) Inf.: Tem rio aqui tomém, o povo gosta muito. Aqui **se fala** muito da Ponte do Coronel. (B, 8, 109)

(53) Inf.: E quanto mais o rio chei, mais mió de **se passar**. (P, 15, 195)

4.4.2 Variáveis independentes

Estabelecemos sete variáveis lingüísticas e três sociais para a investigação dos condicionamentos das variantes da variável dependente discriminadas anteriormente.

4.4.2.1 Variáveis lingüísticas

4.4.2.1.1 Forma antecedente

A variável forma antecedente, atualmente considerada como um caso de “paralelismo formal” nos diversos estudos de fenômenos variados do português (OMENA, 1986; SCHERRE, 1992; CUNHA, 1993; LOPES, 1993; MACHADO, 1995), consiste na repetição, pelo falante, de uma mesma forma numa seqüência discursiva. Ela tem sido observada entre cláusulas, no interior da oração e entre palavras.

Muitas pesquisas de cunho variacionista têm observado a influência da variável paralelismo formal e têm constatado a pertinência de tal influência. Scherre (1992), por

exemplo, ao mencionar diversos estudos que abordaram a influência desta variável, coloca em questão se o efeito paralelístico está associado a um processamento mecânico e superficial ou a funções cognitivas.

De acordo com Lopes (1993), a natureza formal ou funcional parece estar relacionada com o ambiente em que ocorre o fenômeno. Ademais, salienta que no âmbito do discurso parece inegável a influência de outros fatores pragmático-discursivos.

Em alguns trabalhos que analisaram o paralelismo, percebemos que outros fatores condicionam o seu uso. Omena (1986, p. 294) diz que “a probabilidade de se usar *a gente*, ao invés de *nós*, é maior quando o antecedente formal for *a gente* e a referência for igual à anterior. Ela diminui quando se muda a referência”. A referida autora testa o paralelismo juntamente com a referência, sendo assim, quando o falante repete a mesma forma, mantém o mesmo referente, no momento que muda o referente, a forma utilizada será outra.

Cunha (1993, p. 77), ao analisar a variável forma antecedente, questiona: “o processo de repetição acontece independentemente do referente pronominal, isto é, havendo mudança de referente há, conseqüentemente, mudança de forma?”. A autora afirma que a forma antecedente e a referência são interdependentes e que ambas as variáveis, em conjunto, vão definir o processo do paralelismo.

Machado (1995), em seu estudo sobre a variação de *nós* e *a gente*, associa o fator paralelismo ao fator distância entre as orações, objetivando averiguar se há alguma diferença na atuação do paralelismo entre o contexto em que a forma antecedente está mais afastada e aquela em que está mais próxima.

Tendo em vista as leituras realizadas, nossa hipótese é de que ao mencionar uma forma, com o intuito de criar uma harmonia discursiva, o falante manterá a mesma forma e, assim, desencadeará a repetição de formas semelhantes numa série. Postulamos que esta repetição seja influenciada pela referência.

Resolvemos, primeiramente, fazer rodadas isoladas com a variável forma antecedente e, para confirmar que a mudança da forma acompanha a mudança do referente, fizemos também alguns cruzamentos, com o programa CROSSTAB, com os grupos forma antecedente e mudança/manutenção do referente.

Estabelecemos alguns critérios para considerarmos se uma série discursiva é paralela ou não:

- 1) Consideramos como seqüência discursiva aquela em que não ocorre a intervenção do interlocutor na fala do informante (apenas gestos de afirmações – *ham ham*);

- 2) A estrutura analisada não deve estar separada da estrutura anterior a ela equivalente por mais de dez orações, conforme sugere Scherre (1991);
- 3) A estrutura analisada deve fazer parte do mesmo contexto da anterior. Quando o informante muda de assunto no mesmo turno, consideramos a ocorrência como primeira de uma nova série ou forma isolada³², dependendo de a elas se seguirem ou não outras estruturas.

Apresentaremos as possibilidades de forma antecedente na nossa investigação:

- Forma precedida de *nós* explícito

(54) Doc.: Mas... mas tem que tê fé em Deus, né, dona?

Inf.: É... é fé em Deus. Nós tê fé em Deus, que se **Ø num ter** fé em Deus, nada resolvido. Nós tem que pedi a Deus... (P, 7, 153)

- Forma precedida de *nós* implícito

(55) Doc.: Isso que é vontade de ir, né?

Inf.: Nós levava rapadura, carne. Premêro nós levava galinha, **Ø matava** duas galinha, **Ø fritava** e agora botava na mochila, e agora viajava. (P, 13, 167)

- Forma precedida de *a gente* explícito

(56) Inf.: A gente vai mehmo porque **Ø é obrigado ir**, né? (B,13,180)

- Forma precedida de *a gente* implícito

(57) Inf.: Ah! Canjica... a gente pisa o milho, bem pisado, **Ø limpa** ele pele num ficar com pele, com pele. Aí agora, **Ø lava** bem lavado pa tirar aquela tapioca do mi e **Ø põe** no fogo pa cozinhar, né? **Ø põe** pa cozinhar, depois dele cosido hente tempera com que **Ø querer**. **Ø Quiser** ponhar côco, **Ø põe**, se **Ø num quiser**, **Ø põe** leite, açúcar, **Ø tempera** aí. (B, 34, 267)

- Forma precedida de *você* explícito

(58) Inf.: ...você pranta ele, com noventa dia **você** tá colheno ele no pé. (B,13,172)

³² Ignoramos a primeira forma de uma série e a forma isolada por serem estas imunes ao paralelismo. Assim, quando houve uma ocorrência de qualquer uma destas formas, usamos o código não se aplica (/).

- Forma precedida de *você* implícito

(59) Doc.: É a mulher que faz isso aí, né?

Inf.: Põe pra cozinhar e quando cozinhar você vai com um pilão, né? Cê vai, pisa aquele piqui todim pra tirar aquela massa. Aí depois Ø lava de novo, aí agora **Ø pega** aquela massa e Ø põe pra ferver. Aí, cê vai pegano com a cuié. (B,9,144)

- Forma precedida de *eles* explícito

(60) Inf.: Depois que ganhou, ali esquece, qu'eles nem anda aqui. Só **Ø anda** aqui em quatro e quatro ano. (P, 5, 102)

- Forma precedida de *eles* implícito

(61) Inf.: ... Lá em Mato Grosso é verdura, é chuchu, é afrace, é afrace no pé no tempo. É frô, que eles mexe mais lá é com frô. As horta dele, que Ø deve mexer em muita horta de... de verdura assim. **Agora Ø deixou** as verdura, agora é só frô. (B, 10, 148)

- Forma precedida de $\emptyset + V3PS$

(62) Doc.: E o que vocês fazem mais aqui pra se divertir aqui?

Inf.: Aqui?

Doc.: Sim.

Inf.: Aqui num Ø faz muita coisa não. Sempre **Ø faz** festa, trabalho. (B, 7, 70)

- Forma precedida de $\emptyset + V3PP$

(63) Inf.: ... Agora, Ø picaro-lhe fogo, **Ø jogaro** pa debaxo do chão. (P, 12, 183)

- Forma precedida de $\emptyset + VIN F$

(64) Doc.: Eu estou doida pra conhecer essa Ponte do Coronel.

Inf.: Lá é bom, é uma praça assim. É bom de Ø tomar banho, as pedra, uns poço tão bom de **Ø tomar** banho. (B, 4, 54)

- Forma precedida de $\emptyset + V + SE$

(65) Inf.: “É, pra não arrumar namorado”! É o que tou dizeno que eu... eu... aquilo que eu tou lhe falano, que meu tempo foi diferente, porque não se estudava, n' é pra se estudar não, que, n' é pra... pra **Ø aprender**... Mah inda aprendi a... a assinar o nome. (P, 12, 184)

4.4.2.1.2 Mudança/ manutenção do referente

Vários estudos têm mostrado a relevância do fator manutenção ou mudança de referente (OMENA, 1986; CUNHA, 1993; LOPES, 1993; MACHADO, 1995). Este é pautado no pressuposto de que a mudança de referente leva a mudança de forma.

Nos trabalhos analisados, este fator vem conjugado com a variável paralelismo no intuito de observar se, numa seqüência discursiva, a mudança de referente implica alteração na forma de indeterminação.

Neste trabalho, estabelecemos este grupo isoladamente como uma variável, a fim de testar as possibilidades de mudança e manutenção dos recursos de indeterminação em análise numa seqüência discursiva. Seleccionamos os seguintes fatores³³:

- Referente igual à estrutura imediatamente anterior na série discursiva
- Referente diferente da estrutura imediatamente anterior na série discursiva

4.4.2.1.3 Tempo e modo verbal

Buscamos verificar, ao eleger essa variável, se a escolha de uma das variantes está condicionada a uma determinada forma verbal. Para testar a influência desta variável, consideramos a classificação das GTs. Assim sendo, os tempos e modos verbais foram distribuídos da seguinte forma:

- No modo indicativo:
 - presente
 - pretérito perfeito
 - pretérito imperfeito
 - futuro do presente
 - futuro do pretérito
- No modo subjuntivo:
 - presente
 - pretérito imperfeito
 - futuro

³³ Aqui, como fizemos com a variável forma antecedente, não consideramos a primeira forma de uma série e a forma isolada. Assim, quando houve uma ocorrência de qualquer uma destas formas, usamos o código não se aplica (/).

- Formas nominais
 - infinitivo
 - participípio
 - gerúndio

Quando a ocorrência era composta de locução verbal³⁴, consideramos o tempo do verbo auxiliar, como no exemplo (66). Não houve, em nossos dados, ocorrências de locução verbal formada pelos verbos auxiliares ‘ter’ e ‘haver’ que formam tempos compostos.

(66) Inf.: Sei não. Ante da luz era... era quase a merma coisa. Aqui só modificou mermo só a... o que era a coisa que nói num tinha e hoje **tá teno**, né? Hoje **cê pode ter** uma geladeira, **pode ter** um som, **pode ter** qualquer coisa pa fazer parte da lui, você **podeno ter**, ocê tem. (B, 9, 134).

As flexões dos verbos “estar” e “poder”, destacadas nesse exemplo, codificamos como presente do indicativo, com exceção apenas da última ocorrência do verbo “poder” (você **podeno ter**), que codificamos como forma nominal de gerúndio.

Consideramos também a elipse do verbo auxiliar, caso esta elipse fosse de uma forma imediatamente anterior, como em:

(67) Doc.: E tem quanto tempo que tá essa seca aqui assim?
 Inf.: Ah, tá com muitos ano isso, que a gente vem prantano e **Ø perdeno**, moça.
 (P, 5, 94)

4.5.2.1.4 Tipo de oração

Nos trabalhos analisados, há poucas referências sobre o controle da variável tipo de oração sobre o fenômeno da indeterminação. Cunha (1993, p. 47), ao selecionar esse grupo de fatores para análise das formas pronominais *a gente*, *nós* e *você*, diz que “a existência desse grupo baseia-se não em hipótese, mas na necessidade que sentimos de controlar o tipo de oração para mais adiante saber se ele evidencia alguma particularidade do fenômeno”. Ao analisar a variação dos pronomes *nós* e *a gente*, Lopes (1993) e Machado (1995) verificam

³⁴ De acordo com Cunha e Cintra (2001, p. 121),

A locução verbal é o conjunto formado de um verbo auxiliar + um verbo principal, enquanto o último vem sempre numa forma nominal (infinitivo, gerúndio, participípio), o primeiro pode vir:

a) numa forma finita (indicativo, imperativo, subjuntivo): a viticultura **foi-se alargando** talvez a partir do terceiro século.

b) numa forma nominal (infinitivo ou gerúndio): doente, quase não **podendo andar**, fui ter com o evaristo.

uma probabilidade maior para o pronome *nós* em orações independentes e de *a gente* em orações dependentes.

Neste trabalho, consideramos essa variável para verificar se o tipo de oração interfere na escolha das variantes da indeterminação em estudo. Delimitamos, inicialmente, para esta variável, os seguintes tipos de oração:

- Oração absoluta

(68) Doc.: É agitada, como é, muito diferente?

Inf.: **Você tem a festa do mundo todo na cidade.** (B, 9, 130)

- Oração coordenada

(69) Inf.: **Tinha uma igreja véia, derrubaro, fizeram ôtra, tornaro a derrubá, agora fizeram aquela.** Agora ali tá feita. (P, 7, 133)

- Oração principal

(70) Doc.: É muito grande, né?

Inf.: **A gente pensa** que a gente tá passano na mesma rua que a gente entrou, pois não entra, passa ni outra. (B, 10, 166)

- Oração subordinada

(71) Inf.: ... Tão matano **como mataro um fio meu.** (P, 12, 182)

Observamos que existem determinadas orações que podem ser classificadas como principal ou subordinada e ao mesmo tempo coordenada. Devido a isso, acrescentamos mais dois fatores que dessem conta desses casos:

- Oração principal composta de duas coordenadas

(72) Inf.: Aí depois que vai sortano aquele óleo, **a gente vai panhano e ponhano na vasila.** (B, 7, 80)

- Oração subordinada composta de duas coordenadas

(73) Doc.: Para vocês parece que todo mundo sabe, mas na verdade sabe não.

Inf.: É, perfeitamente. A gente sabe o ritmo **que a gente nasceu e criou**. (B, 13, 171)

4.5.2.1.5 Tipo de verbo

O controle desta variável busca verificar se as variantes da indeterminação sofrem alguma influência do tipo de verbo da oração em que ocorrem. Na análise desse grupo de fatores, classificamos os verbos de acordo com a GT. Selecionamos três tipos:

- Verbo de ligação (cópula)

(74) Doc.: Só é dona de casa, cuidano dos filho... mas também...

Inf.: Às vezes num tem nada pa hente fazer. Num tem licuri, num tem mamona, a hente só **Ø fica** dento de casa com braço encruzado. (P, 2, 37)

- Verbo transitivo

(75) Doc.: Como é que faz aquilo?

Inf.: Ali é... **Ø cavaca** a terra, móia, ma... **Ø móia** bem moiadim e **Ø massa** o barro bem massadinho mó dele ficar assim liguento. (B, 7, 80)

- Verbo intransitivo

(76) Doc.: A senhora gosta de cozinhar?

Inf.: Gosto de cozinhar. A vida da gente é essa mesmo.

Doc.: É.

Inf.: Tem que **Ø trabalhar**, **Ø cozinhar**, cuidar da casa, né não? [rindo]. (B, 10, 150)

Consideramos como transitivos os verbos transitivos diretos, indiretos, bitransitivos, transitivos relativos e transitivos circunstanciais.

Na análise dos verbos transitivos e intransitivos, partimos do que foi proposto por Perini (1996, p. 164) que diz que “a transitividade deve ser feita em termos de exigência, recusa e aceitação livre de cada uma das funções relevantes”. Para isso, é necessário que na análise da transitividade levemos em consideração o contexto, pois esta é feita de acordo com

o texto, e não isoladamente. Assim, o mesmo verbo pode estar empregado tanto intransitivamente, como transitivamente.

4.5.2.1.6 Preenchimento do sujeito

No nosso trabalho, selecionamos não só os casos de presença dos pronomes sujeito, mas também aqueles com sujeito implícito (oculto). A inclusão deste grupo de fatores constitui um meio de controlar a ocorrência de sujeito preenchido ou não.

Os fatores estabelecidos foram os seguintes:

- Sujeito pronominal explícito: quando o pronome se faz presente na ocorrência analisada.

(77) Doc.: Pra dor de cabeça assim, dor de barriga tem alguma folha assim, que vocês usam aqui? (...)

Inf.: Pra dor de cabeça e pra dor de barriga, **a gente** tem um remédio assim pra dor de barriga que **a gente** faz, **a gente** tem a tapioca de araruta... (B, 8, 108)

- Sujeito pronominal implícito: quando o pronome não se faz presente na ocorrência analisada.

(78) Inf.: Da rapadura! Ali cê vai controlano o fogo ali... e... \emptyset vai jogano massa. Aí ela vai enxugano, enxugano... (B, 9, 144)

Na análise desta variável, codificamos apenas os recursos de indeterminação representados por pronomes, no nosso caso, os indeterminadores *nós*, *a gente*, *ocê* e *eles*. Os casos dos recursos representados por verbos com sujeito vazio, como $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF$ e $\emptyset+V+SE$, receberam o código não se aplica (/).

4.5.2.1.7 Grau de indeterminação

Já vimos no item 3.3.3 que o conteúdo genérico de cada recurso de indeterminação tem características referenciais diferenciadas e, em alguns casos, podem indicar inclusão ou exclusão das pessoas do discurso.

Para análise desta variável, definimos, seguindo o que foi proposto por Cunha (1993), três graus de indeterminação de acordo com a referência ao contexto:

- Indeterminação completa – quando não há possibilidade de recuperação do referente dentro do contexto.

(79) Doc.: E o senhor já foi ofendido?

Inf.: Não senhora, nunca não. Só de... desse bichinho que **Ø chama** la... lacraia.

Doc.: Ham.

Inf.: Que **Ø chama** candradão, isso aí já me ofendeu já. (P, 5, 100)

- Indeterminação parcial com referência implícita no contexto – quando o referente não está presente no texto, mas pode ser depreendido do contexto através de inferência.

(80) Doc.: E vem cá, aqui tem muito morro?

Inf.: Possui bastante.

Doc.: Você já subiu em algum?

Inf.: Já. Eu já e muito. Quano **a gente** vai assim tirar casca, a gente sobe num morro arto. (B, 4, 66)

- Indeterminação parcial com referência explícita no contexto – quando há uma relação com um determinado referente possibilitando sua interpretação, sendo este recuperado por meio de inferência.

(81) Inf.: Mai no tempo que **a gente** era rapaz, que **a gente** juntava o grupo da gente todim... ah! O dia de hoje, nós tava hoje aqui? O forró tava aí por conta! (B, 9, 140)

4.5.3.2 Variáveis extralingüísticas

4.4.3.2.1 Gênero³⁵

O gênero tem sido freqüentemente observado nos estudo variacionistas. Esses estudos apontam a diferença existente no comportamento lingüístico de homens e mulheres. De acordo com Trudgill (1974, apud SILVA; PAIVA, 1998, p. 355),

³⁵ Optamos por utilizar o termo gênero ao invés de sexo, pois de acordo com estudos na área de antropologia e sociologia, este último é biológico, inato e imutável. Já o gênero é o sexo socialmente construído, podendo ser modificado. De acordo com Rago (1998, p. 90) o gênero é a “construção social e cultural das diferenças sexuais”. Esse termo surge para explicar o que a diferença sexual, biológica não consegue explicar. No âmbito da sociolingüística, os estudos têm mostrado que o comportamento lingüístico de homens e mulheres não se explica a partir das diferenças biológicas, mas também dos comportamentos sociais, de classe etc.

geralmente, a correlação sexo/variação mostra que as mulheres tendem a preferir formas socialmente valorizadas, tendência essa freqüente em situação de variação estável. Uma inversão desta tendência pode também ser tomada com indicação de que uma nova forma está se implementando na língua.

Uma das características que tem sido apontada nos estudos é que as mulheres têm maior preferência pelas variantes lingüísticas mais prestigiadas socialmente, demonstrando uma maior consciência por parte delas do *status* social das formas lingüísticas.

O papel da variante gênero na mudança lingüística é muito importante. Porém, como foi dito na citação acima, um aspecto que se deve levar em consideração é o valor social da variante inovadora. Assim, quando se trata de implementar na língua uma forma socialmente prestigiada, as mulheres tendem a liderar processos de mudança (LABOV, 1966). Ao contrário, quando se trata de implementar uma forma socialmente desprestigiada, as mulheres assumem uma atitude conservadora e os homens tomam a liderança do processo.

Com relação aos hábitos lingüísticos, a mulher até poderia ser considerada conservadora, mas isso não significa que sua linguagem seja estática. Ela adere às inovações que demonstram uma aceitabilidade social. Em muitos processos de mudança não está envolvida uma polarização evidente entre uma variante de prestígio e uma variante não prestigiada.

Conforme Paiva (2003, p. 37), “o padrão de comportamento da variável sexo procede de uma generalização que não corresponde inteiramente à realidade, pois está baseado em análises dessa variável isoladamente”. Segundo ela, do cruzamento desta variável com outras, como classe social, idade, estilo de fala, escolaridade, “podem emergir padrões de correlação diferenciados que apontam a relatividade das correlações entre uso de variantes lingüísticas e o sexo do falante”. De acordo com Labov (1983), no estágio atual das pesquisas, a maioria de nossas explicações é apenas especulação, porque os padrões de correlação não são fixos e inteiramente consistentes.

Em nosso trabalho, não temos uma hipótese bem específica para a variante gênero na indeterminação do sujeito, mas esperamos encontrar diferenças entre o comportamento dos homens e das mulheres frente ao fenômeno, quer na freqüência com que o empregam, quer na escolha de um ou de outro recurso que o expresse.

4.4.3.2.2 Faixa etária

As pesquisas sociolinguísticas têm demonstrado a importância do exame do comportamento linguístico por diferentes faixas etárias a fim de comprovar a existência de variantes conservadoras e inovadoras, bem como indicar se o fenômeno em estudo encontra-se em um processo de variação estável ou mudança linguística.

Ao examinar a relação da mudança linguística com a idade do falante, Naro (2003, p. 44) afirma que

os falantes adultos tendem a preferir as formas antigas, criando uma situação estranha, pelo menos à primeira vista: existem pessoas que, apesar de estarem em interação constante (do tipo pai/filho), costumam falar de maneira distinta. Entretanto, isso não chega a comprometer a comunicação, já que ambos os lados são capazes de utilizar e entender todas as formas. Trata-se apenas de uma tendência em direção a outra forma. Com o correr do tempo, é provável que a forma nova seja adotada por todos.

No caso das comunidades em estudo neste trabalho, é necessário recorrer a essa variável para que se possa verificar se há e quais são as diferenças no uso das variantes para a indeterminação do sujeito por parte de falantes pertencentes a faixas etárias diferentes. Desta forma, poderemos observar se certas variantes têm a preferência dos jovens ou se são mais usadas por pessoas mais velhas.

Seguindo o modelo da pesquisa sociolinguística, os dados analisados nesta pesquisa foram distribuídos em três faixas etárias:

- Faixa 1 – 20 a 35 anos
- Faixa 2 – 36 a 55 anos
- Faixa 3 – acima de 56 anos

4.4.3.2.3 Localidade

Entre os grupos de fatores extralinguísticos tradicionalmente considerados em pesquisas de cunho dialetológico ou sociolinguístico, costuma-se inserir a origem dos locutores para efetivar-se o controle da variação diatópica. Este controle baseia-se na premissa de que o comportamento linguístico de falantes de um mesmo idioma pode diferenciar-se de acordo com o maior ou menor grau de isolamento que ocupam e que determinam situações sociais específicas.

De acordo com Machado (1995, p. 136), “os pontos de inquérito podem apresentar comportamento lingüístico diferenciado em decorrência das características fisiográficas, socioeconômicas, históricas e culturais de cada comunidade em questão”. Supomos que a influência da variável diatópica no uso das formas de indeterminação do sujeito possam ser explicadas levando-se em conta: a) a questão da proximidade *versus* distanciamento dos centros urbanos e, conseqüentemente, a do maior ou menor isolamento; e b) a questão do povoamento mais recente *versus* povoamento mais antigo dessas regiões.

Assim, buscaremos analisar o condicionamento da variável localidade no falar dos moradores de Bananal/ Barra dos Negros e de Piabas.

5 ANÁLISE DOS DADOS

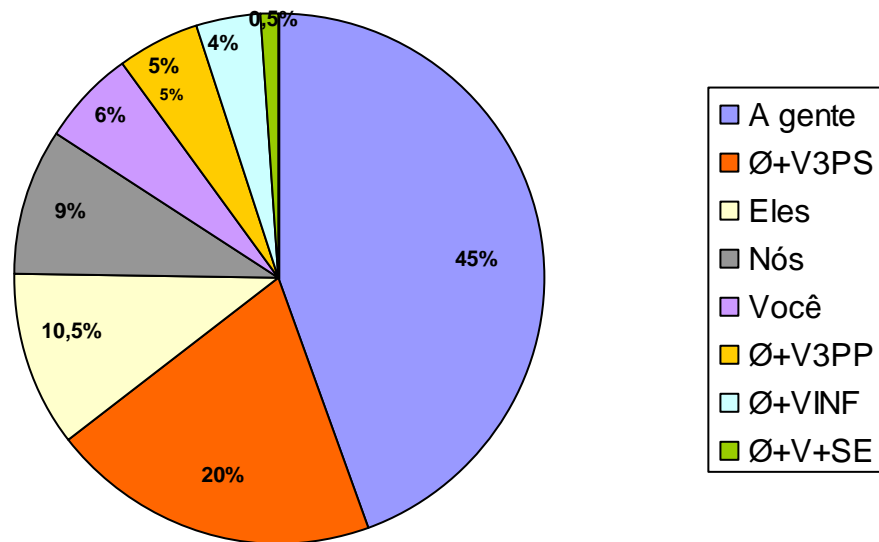
Neste capítulo, vamos expor a análise dos resultados numéricos obtidos através do processamento dos dados feito pelo pacote de programas VARBRUL. Consideramos, em tabelas e gráficos, os resultados apontados pelos programas MAKECEL, CROSSTAB e VARB2000. O total geral de foi de 2.569 ocorrências retiradas de 24 entrevistas do tipo DID. Para verificar de que modo se distribuem nos *corpora* os recursos de indeterminação do sujeito, *a gente*, *nós*, *você*, *eles*, $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF$ e $\emptyset+V+SE$, calculamos suas frequências e as expomos, em ordem decrescente, na tabela 1:

Tabela 1: Frequência dos recursos de indeterminação do sujeito nos *corpora*.

VARIANTES	Nº	%
A gente	1.149	45
$\emptyset+V3PS$	525	20
Eles	272	10,5
Nós	227	9
Você	151	6
$\emptyset+V3PP$	120	5
$\emptyset+VINF$	109	4
$\emptyset+V+SE$	15	0,5
TOTAL	2569	100

Distribuímos as variantes da indeterminação em estudo no gráfico 1, para melhor visualização:

Gráfico 1: Distribuição dos recursos de indeterminação do sujeito nos *corpora*.



Conforme já esperávamos, os dados mostram que os falantes das duas comunidades analisadas usam outros recursos para indeterminar o sujeito além dos citados pela GT: $\emptyset+V3PP$ e $\emptyset+V+SE$. Além disso, esses dois recursos não se mostraram muito produtivos na amostra se compararmos com as outras formas utilizadas; as duas juntas não chegam nem a 6% de uso no total. Os dados evidenciam uma alta produtividade dos pronomes pessoais, principalmente a forma *a gente*, como recurso de indeterminação, constituindo a maioria das ocorrências: 70% do total geral dos dados contra 30% de formas sem sujeito lexical. Neste último, é a forma $\emptyset+V3PS$ a mais produtiva.

Um outro fato interessante a ser observado é a relação entre o uso de *nós* e *a gente*. Percebemos uma redução no percentual de uso da forma *nós*: enquanto o uso da forma *a gente* equivale a quase metade do total de ocorrências dos *corpora*, a forma *nós* equivale a apenas 1/10. Nossa hipótese para essa queda no uso de *nós* pode estar relacionada à redução nos paradigmas flexionais do português do Brasil (cf. MONTEIRO, 1994; DUARTE, 1995; MENON, 1995). DUARTE (1995) mostra o reflexo dessa redução de *nós* com referência determinada e sua substituição crescente por *a gente*. Podemos considerar, desse modo, que essa diminuição no uso de *nós* determinado esteja se refletindo em seu uso indeterminado.

Em relação ao total de ocorrências, o gráfico 1 revela que o recurso mais utilizado como forma de indeterminação do sujeito no português rural das comunidades analisadas é *a*

gente (45%). Ele vem seguido pela forma de $\emptyset+V3PS$ (20%), da forma *eles* (10,5%) e da forma *nós* (9%). As formas $\emptyset+V3PP$, *você*, $\emptyset+VINFIN$ e $\emptyset+V+SE$ tiveram as menores frequências, 6%, 5%, 4% e 0,5%, respectivamente.

Ao confrontarmos os nossos resultados com os de outros estudos, com análise de outros *corpora* sobre a indeterminação, vimos que eles divergem. A maioria das pesquisas confirma a prevalência do uso da forma *você*, para indeterminar o sujeito, sobre as demais formas. Rollemberg (1991), ao tratar da norma urbana culta de Salvador, encontra 35,54% de uso desta forma; do mesmo modo, Cunha (1993), analisando a norma urbana culta do Rio de Janeiro, encontra 48% de preferência de uso pelos falantes. Cavalcante (1998), tratando do uso de diversas formas de indeterminação no português culto de Fortaleza, encontra 29% de uso do *você*, sendo este também o recurso mais usado. Em Godoy (1999), na análise de 72 entrevistas que compõem o *corpus* do Projeto VARSUL, *você* é também o indeterminador mais usado, com 37% do total de ocorrências. Apenas nos trabalhos de Almeida (1992), que analisa os recursos de indeterminação presentes na amostra do Projeto Censo Variação – RJ, e Milanez (1982), que analisa a norma culta de São Paulo, a forma *a gente* é a mais utilizada pelos usuários da língua.

Apesar de o objetivo deste trabalho ser unir comunidades que compartilhem fatos sociohistóricos, ampliando a representação desse dialeto em diversas regiões do estado da Bahia, buscamos verificar o comportamento das variantes em cada comunidade isoladamente a fim de observar a existência de alguma particularidade entre elas. Do total geral de ocorrências (2.569), 1.775 (69%) ocorrem na fala dos informantes de Bananal / Barra dos Negros, enquanto 794 (31%) ocorrem na fala dos informantes de Piabas. A frequência das variantes em cada comunidade, obtida pelo MAKECELL, é apresentada na tabela a seguir:

Tabela 2: Distribuição dos dados de acordo com as comunidades.

Localidade/ Variantes	Bananal / Barra dos Negros		Piabas	
	Nº	%	Nº	%
A gente	881	50	268	34
Nós	125	7	102	13
Você	139	8	13	2
Eles	186	10	86	11
Ø+V3PS	338	19	185	23
Ø+V3PP	37	2	83	10
Ø+VINF	64	4	47	6
Ø+V+SE	5	0	10	1
TOTAL	1775	100	794	100

Analisando isoladamente as comunidades de Bananal/ Barra dos Negros e Piabas, verificamos que as formas mais utilizadas em ambas as comunidades são as mesmas: *a gente* e \emptyset +V3PS, com um percentual de uso de 50% e 19%, respectivamente, em Bananal / Barra dos Negros e de 34% e 23%, respectivamente, em Piabas.

Ao confrontar as comunidades, observamos que as variantes *a gente* e *você* têm um índice de uso maior entre os falantes de Bananal / Barra dos Negros do que entre os de Piabas. Por outro lado, as variantes *nós*, \emptyset +V3PS, \emptyset +V3PP, \emptyset +VINF e \emptyset +V+SE têm uma frequência de uso maior em Piabas do que em Bananal / Barra dos Negros. A variante *eles* é usada de modo uniforme em ambas as comunidades.

Vemos, assim, que em Piabas há uma tendência evidente de uso das formas consideradas “conservadoras” como as formas \emptyset +V3PP (10% de uso em Piabas, contra 2% em Bananal / Barra dos Negros) e \emptyset +V+SE. Esta última, em relação ao percentual, não apresenta grandes diferenças devido à pequena quantidade de dados, porém em números absolutos a comunidade de Piabas apresenta o dobro do uso de Bananal / Barra dos Negros. Observamos, ainda, um uso expressivo da forma *nós* que, apesar de não ser apresentada na GT como forma para indeterminar o sujeito, é considerada conservadora enquanto pronome.

Assim exposto, analisaremos, nas próximas seções, os fatores que condicionam o uso de cada uma dessas variantes. Antes de iniciarmos nossa análise, esclarecemos que, para efetuar as rodadas estatísticas, tivemos que excluir as ocorrências de \emptyset +V+SE devido ao

pequeno número de dados (apenas 15), o que ocasionou um grande número de *knockouts*³⁶, impossibilitando a aplicação da regra variável. Passamos, então, a trabalhar com 2.554 ocorrências.

De acordo com as possibilidades oferecidas pelo VARBRUL, podem ser realizadas rodadas binárias (com duas variantes), ternárias (com três variantes) e eneárias³⁷ (com quatro ou cinco variantes). Por termos sete variantes da variável dependente, partimos para a observação de cada variante em confronto com as demais, fazendo, desta forma, rodadas binárias. Nesta fase, o programa seleciona os grupos de fatores considerados como favorecedores para aplicação da regra e descarta os que não a favorecem. Analisaremos cada variável separadamente, observando: 1) a descrição das modificações feitas em alguns grupos de fatores devido à não ocorrência ou pequeno número de dados; 2) a apresentação dos grupos considerados como favorecedores para cada uma, na ordem de seleção oferecida pelo programa; e 3) a interpretação lingüística dos resultados numéricos.

5.1 VARIÁVEIS SIGNIFICATIVAS SELECIONADAS PELAS FORMAS DE INDETERMINAÇÃO DO SUJEITO

5.1.1 A gente

Na primeira rodada dos dados, que teve como valor de aplicação a variante *a gente*, o programa excluiu como não favorecedores a aplicação da regra os grupos tipo de oração, tipo de verbo, preenchimento do sujeito e mudança/ manutenção do referente. Excluímos esses grupos de fatores do arquivo de condição e na segunda rodada fizemos ainda algumas modificações: 1) retiramos os tempos e modos verbais futuro do presente do indicativo e presente e pretérito imperfeito do subjuntivo, por apresentarem pouco ou nenhum dado; e 2) amalgamamos todas as formas pronominais de sujeito explícito com sujeito implícito, além de amalgamarmos as formas com sujeito vazio ($\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINFL$ e $\emptyset+V+SE$) em um único fator, no grupo forma antecedente, também devido ao pequeno número de dados.

³⁶ *Knockout* é o termo usado, no VARBRUL, para indicar que um determinado resultado é absoluto, ou seja, ou a regra não se aplica e esse resultado é 0%, ou todos os dados indicam a aplicação da regra e o resultado é 100% e, então, não há variação para analisar. Para que o programa prossiga na execução das rodadas estatísticas, é necessário que se retire da análise o fator com o qual se dá o *knockout*.

³⁷ Cf. o item 4.2: O suporte quantitativo.

Apresentadas as modificações feitas, veremos agora os grupos considerados como favorecedores para o pronome *a gente* como recurso de indeterminação do sujeito:

- 1° Forma antecedente
- 2° Grau de indeterminação
- 3° Tempo e modo verbal
- 4° Localidade
- 5° Faixa etária
- 6° Gênero

5.1.1.1 Forma antecedente

A forma antecedente mostrou-se, em nossa investigação, o fator mais significativo, sendo selecionado na primeira ou na segunda posição em todas as rodadas estatísticas. A hipótese que originou esse grupo foi a mesma proposta por Scherre e Naro (1993, p. 2): a de que “marcas levam a marcas”. Assim, a opção por uma determinada forma no início da seqüência discursiva tende a desencadear uma série de repetições da forma empregada na primeira referência. Vejamos os resultados da tabela 3:

Tabela 3: Influência da forma antecedente no uso de *a gente*.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	622/674	92	0,93
Nós	15/122	12	0,12
Você	5/107	5	0,06
Eles	19/160	12	0,12
Formas com sujeito zero	21/351	6	0,08
TOTAL	682/1414	48	-

Nível de significância = 0,000

De acordo com os resultados desta tabela, a maior probabilidade de ocorrência da forma *a gente* está nos enunciados que têm como forma precedente também *a gente* (explícito ou não), tendo como peso relativo 0,93. Observando os resultados de Cunha (1993), vemos que este também é o primeiro grupo de fatores selecionado e que a forma *a gente* tende a ocorrer depois de outro *a gente* que o precede no discurso.

Ao analisar nossos resultados, questionamos se haveria fatores que pudessem intervir no processo de repetição do antecedente formal, pois, de acordo com Omena (1986, p. 295), “a escolha de uma das formas, ao se nomear pela primeira vez um referente na seqüência de um discurso, deve-se a outros fatores”. Omena testa o paralelismo juntamente com a referência, pois, segundo ela, a mudança da forma está condicionada à troca de referência. A partir do que foi dito por Omena e da observação dos nossos dados vimos que poderia haver a interferência da variável mudança/ manutenção do referente, apesar de este grupo de fatores não ter sido selecionado pelo programa em nenhuma rodada. Acreditando que com a mudança de referente há uma mudança na forma utilizada pelo falante, fizemos um cruzamento com os grupos forma antecedente X mudança/ manutenção do referente, através do programa CROSSTAB, e tivemos o seguinte resultado:

Tabela 4: Efeito da forma antecedente e da mudança/ manutenção do referente no uso de *a gente*.

Referência/ Forma antecedente	Igual		Diferente	
	Nº/T	%	Nº/T	%
A gente	617/634	97	5/40	13
Nós	15/117	13	0/10	0
Você	5/103	5	0/4	0
Eles	0/134	0	19/26	73
Formas com sujeito zero	7/302	2	14/49	29

Os resultado da tabela 4 mostram que a manutenção do referente acompanha a manutenção da forma de indeterminação escolhida pelo falante. Assim, quando o falante faz referência a um sujeito usando a forma indeterminada *a gente*, em 97% dos casos ele mantém a mesma forma na seqüência discursiva. Quando há a mudança de referência, o falante se vê diante de uma situação de escolha de uma nova forma, mantendo a mesma forma em apenas 13% dos casos. Ao analisar nos *corpora* esses 13% de casos de mudança de referência, mas mantendo a mesma forma, o que corresponde a 5 ocorrências, vimos que eles estão presentes na fala de uma única informante de Bananal / Barra dos Negros da faixa etária 2. Vejamos um trecho de sua fala no exemplo (82):

- (82) Doc.: E a senhora vai sempre em Mato Grosso? (...)
 Inf.: De vez em quando eu vou lá, mas é ni ocasião de festa, né? (...)
 Doc.: (...) a senhora vai na... de carro também, quando vai?
 Inf.: Eu vou de carro, e no tempo aqui atrás, que num tinha estrada que ia pra lá, quonde a gente era tudo mocinha nova, que tem muita vontade de andar, né?
 Doc.: Ham.
 Inf.: **A gente** ia de pé, mas agora ninguém tá ino de pé mais não.
 Doc.: Hum.
 Inf.: **A gente** tá ino mais de carro. (B, 12, 212-3)

Essa seqüência discursiva ressalta o uso diferenciado da forma de indeterminação *a gente* pela informante. Ao falar de que forma vai a Mato Grosso (comunidade vizinha a de Bananal / Barra dos Negros), a informante faz referência a duas épocas: uma anterior àquela dos interlocutores do diálogo, quando a informante ainda era “mocinha nova”; e outra, ao momento em que a entrevista está sendo realizada, em que a informante está mais velha. No primeiro caso, *a gente* é usada pela informante com referência a ela e as moças jovens e no segundo caso, a forma *a gente* é usada com referência a ela mais qualquer pessoa da comunidade, não especificando quem são essas pessoas, se são as mais jovens ou se são as mais velhas. A mudança de referência com manutenção da mesma forma foi também um fenômeno constatado por Cunha (1993).

Ainda de acordo com os dados da tabela 4, observamos que no momento em que o falante muda a referência ocorre com maior freqüência a forma *eles* (73%). Fizemos uma análise nos *corpora* para verificar em que contextos ocorrem as 19 ocorrências de *eles* após a forma *a gente* e vimos que os informantes mudam de forma quando no mesmo contexto discursivo eles fazem referência a dois grupos distintos e em um deles o próprio informante não está inserido. O exemplo (83) mostra esse caso:

- (83) Inf.: (...) E aí agora **a gente** tá aqui na área do DENOC. Onde é que **a gente** quase nem já pode construir uma casinha, divido **da gente** ficou muito esfracassado depois que eles fez essa barragem, né? **A gente** num prantou roça... (B, 12, 227)

Nesse exemplo, quando o informante fala das dificuldades que tem para construir uma casa para morar, devido à falta de condição financeira e, conseqüentemente, de trabalho, ele usa *a gente* para se referir a ele e às pessoas da comunidade e, no momento que ele deixa de se referir às pessoas de sua comunidade e passa a se referir às pessoas que trabalham no DENOC, ele muda também a forma utilizada: ao invés de *a gente*, passa a usar *eles*. Esses

casos são freqüentes nos dados, nas 19 ocorrências de uso de *eles* após *a gente*, os informantes mudam de forma quando se referem a grupos distintos, como as pessoas que trabalham no DENOC (conforme foi mostrado no exemplo acima), os políticos, as pessoas que moram na cidade de Rio de Contas ou na comunidade de Mato Grosso e as pessoas que trabalham no posto de saúde local.

5.1.1.2 Grau de indeterminação

Esta foi a segunda variável selecionada pelo programa VARB2000. De acordo com os resultados da tabela 5, vemos que o falante usa a forma *a gente* quando a indeterminação é parcial com referência implícita no contexto, com peso relativo 0,65.

Tabela 5: Influência do grau de indeterminação no uso de *a gente*.

Grupo: Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	370/1132	33	0,38
Indeterminação parcial com referência implícita no contexto	727/1194	61	0,65
Indeterminação parcial com referência explícita no contexto	52/228	23	0,31
TOTAL	1149/2554	45	-

Nível de significância = 0,000

Esse tipo de indeterminação é o que mais ocorre nos *corpora*, seu uso com a forma *a gente* se dá quando o falante faz referência a ele mesmo e a outras pessoas, excluindo o interlocutor e sem discriminar quem são essas pessoas. O falante não especifica quem é, mas pelo contexto podemos deprender o referente. É o que acontece em (84):

(84) Doc.: Quando é que começa a estação de chuva aqui?

Inf.: A estação de chuva aqui?

Doc.: Ham.

Inf.: No mês de mauço, que é o mês que a gen... **a gente** chama aqui mês da quaresma, né? (B, 3, 19-20)

Nesse exemplo, observamos que, ao usar a forma *a gente*, o falante faz referência a ele mais às pessoas que vivem na comunidade de modo generalizado, sendo esta referência depreendida a partir do contexto por meio de inferência. Nesse exemplo, o uso do advérbio ‘aqui’ também facilitou para que identificássemos o referente.

5.1.1.3 Tempo e modo verbal

Este foi o terceiro grupo de fatores selecionado pelo VARBRUL como significativo para aplicação da variante em estudo. Os índices percentuais e probabilísticos encontram-se na tabela 6:

Tabela 6: Influência do tempo e modo verbal no uso de *a gente*.

Grupo: Tempo e Modo Verbal	Apl/T	%	P.R.
Presente do indicativo	817/1630	50	0,55
Pretérito perfeito do indicativo	45/237	19	0,24
Pretérito imperfeito do indicativo	130/321	40	0,50
Pretérito imperfeito do subjuntivo	3/8	38	0,21
Futuro do subjuntivo	20/55	36	0,35
Infinitivo	113/274	41	0,41
Gerúndio	18/23	78	0,92
TOTAL	1146/2548	45	-

Nível de significância = 0,000

De modo geral, os resultados mostram que as maiores probabilidades para o uso de *a gente* ocorrem com o gerúndio (0,92) e com o presente do indicativo (0,55), como foi encontrado também por Godoy (1999).

Ao analisarmos as ocorrências com gerúndio, presentes nos *corpora*, percebemos que os falantes utilizam esta forma em substituição as do futuro do subjuntivo, em orações subordinadas adverbiais condicionais, atribuindo ao discurso uma característica hipotética, o que é propício para a indeterminação, como podemos ver nos exemplos (85a) e (85b):

(85a) Inf.: Aqui dia de feira **a gente quereno ver piega** é só mandar tocar um violão e chamar uns... encosta uns moleque aí e n'estante ver a pera avoar. (P, 15, 197)

(85b) Inf.: Aqui dia de feira **se a gente quiser ver piega** é só mandar tocar um violão e chamar uns... encosta uns moleque aí e n'estante ver a pera avoar. (P, 15, 197)

Observamos que o uso do presente do indicativo com a forma *a gente* se dá, porque, na maioria dos contextos, o informante usa *a gente* para descrever ou explicar ações habituais praticadas com frequência sem fazer referência à sua duração, o que caracteriza esse tempo.

Vejam os exemplos de uso da forma *a gente*, retirados dos *corpora*, com o verbo flexionado no presente do indicativo:

- (86) Doc.: Ah! E a tapioca, a senhora sabe fazer?
 Inf.: Ah, sei! A tapioca...
 Doc.: Então me ensina [rindo].
 Inf.: [rindo] Óia, a tapioca, **a gente tem** que relar a... a mandioca e \emptyset prontar a massa. \emptyset **Pega** a massa, **a gente bota** um pouco de água na massa. Aí agora, \emptyset **joga** no saco ó, \emptyset **joga** no saco pra sair a tapioca.
 Doc.: Hum.
 Inf.: \emptyset **Faz** com aquela água a tapioca, depois que assenta água, ó, aí agora, **a gente escorre** a água e \emptyset **bota** a tapioca pra enxugar. Depois seca... seca, aí **a gente ensaca**. (B, 12, 207).

Um outro aspecto observado foi que a forma *a gente* ocorreu sempre com a forma verbal com morfologia zero de número e de pessoa. Como em:

- (87a) Inf.: [rindo] Óia, a tapioca, **a gente tem** que relar a... a mandioca...
 e não
 (87b) Inf.: [rindo] Óia, a tapioca, **a gente temos** que relar a... a mandioca...

5.1.1.4 Localidade

A localidade foi o quarto grupo selecionado pelo programa. Os falantes das localidades pesquisadas apresentam comportamento linguístico diferenciado, conforme observamos na tabela 7:

Tabela 7: Influência da localidade no uso de *a gente*.

Grupo: Localidade	Apl/T	%	P.R.
Bananal/Barra dos Negros	881/1765	50	0,55
Piabas	268/774	34	0,39
TOTAL	1149/2554	45	-

Nível de significância = 0,000

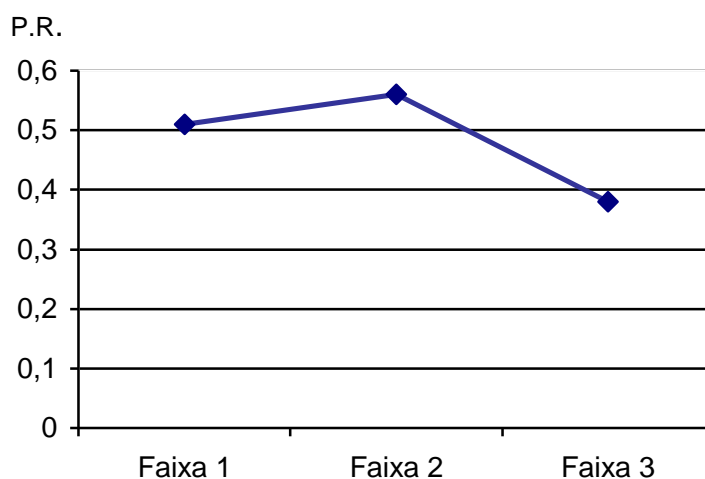
É na comunidade de Bananal / Barra dos Negros que os falantes utilizam mais a forma *a gente* para indeterminar o sujeito, tendo um peso relativo 0,55. Se compararmos esse resultado com o da comunidade de Piabas, veremos que esta última tem um peso relativo bem abaixo (0,39). Acreditamos que o maior uso de *a gente* nas comunidades de Bananal / Barra

dos Negros seja devido ao turismo que há na região onde elas se localizam, recebendo, assim, influência de outros dialetos, o que pode estar influenciando o uso desta forma pelos moradores.

5.1.1.5 Faixa etária

O quarto fator considerado como significativo para o uso da forma *a gente* foi a faixa etária do informante.

Gráfico 2: Influência da faixa etária no uso de *a gente*.



Nível de significância = 0,000

Os dados do gráfico 2 apontam a preferência pela forma *a gente*, para indeterminar o sujeito, entre os falantes da faixa 2, com um peso relativo 0,56. Na faixa 1, a probabilidade do uso desta forma é neutra (0,51). Os falantes da faixa 3 registram o índice probabilístico mais baixo (0,38) de emprego desta forma, o que nos mostra ser esta uma variante inovadora também para indeterminar o sujeito.

Foi necessário conjugar esse grupo de fatores com o grupo localidade para sabermos de forma mais específica quem são esses informantes da faixa 2, visto que as comunidades em estudo têm comportamentos lingüísticos diferenciados.

Tabela 8: Efeito da faixa etária e da localidade no uso de *a gente*.

Faixa Etária/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Faixa 1	323/647	50	0,53	132/335	39	0,45
Faixa 2	380/681	56	0,63	113/298	38	0,42
Faixa 3	178/442	40	0,44	23/151	15	0,23
TOTAL	881/1770	50	-	268/784	34	-

Nível de significância = 0,000

A partir dos resultados desta tabela, vemos que os informantes da faixa 2 de Bananal / Barra dos Negros são os que mais utilizam a forma *a gente*, em frequência e em peso relativo, 56% e 0,63, respectivamente. A faixa 1 tem um uso reduzido desta forma, porém não deixa de ser significativa, pois o peso relativo é 0,53, e na faixa 3 o uso de *a gente* é ainda mais reduzido, sendo insignificante para o programa (0,44). Já em Piabas, o uso de *a gente* não é significativo, apresentando peso relativo menor que 0,50 em todas as faixas etárias. Porém, é interessante observar que nesta comunidade há um uso crescente desta variante na medida que a idade do informante diminui: 0,23 na faixa 1, 0,42 na faixa 2 e 0,45 na faixa 3.

Vimos, então, que são os informantes da faixa 2 seguidos dos da faixa 1 que usam mais significativamente a forma *a gente*. Esses resultados, no entanto, não estão em consonância com os resultados de outros trabalhos sobre a inserção da forma *a gente* no sistema pronominal do português do Brasil, os quais evidenciam uma tendência ascendente ao uso de *a gente*, tanto em dialetos populares quanto em dialetos cultos, principalmente entre os falantes da geração mais nova (cf. FERNANDES; GORSKI, 1986; OMENA, 1986; ALBÁN; FREITAS, 1991b; MACHADO, 1995). Por outro lado, nos trabalhos de Almeida (1992) e de Cunha (1993), que analisam as formas de indeterminação do sujeito na fala carioca, os resultados são simétricos aos nossos: é na faixa 2 que há o uso mais freqüente da forma *a gente*. Nos trabalhos de Setti (1997) e Godoy (1999), que analisam amostras do *corpus* do VARSUL, a maior probabilidade de uso desta variante está na faixa 3.

5.1.1.6 Gênero

O gênero do informante foi o último grupo selecionado na análise da variante *a gente*. Os resultados encontram-se na tabela 9:

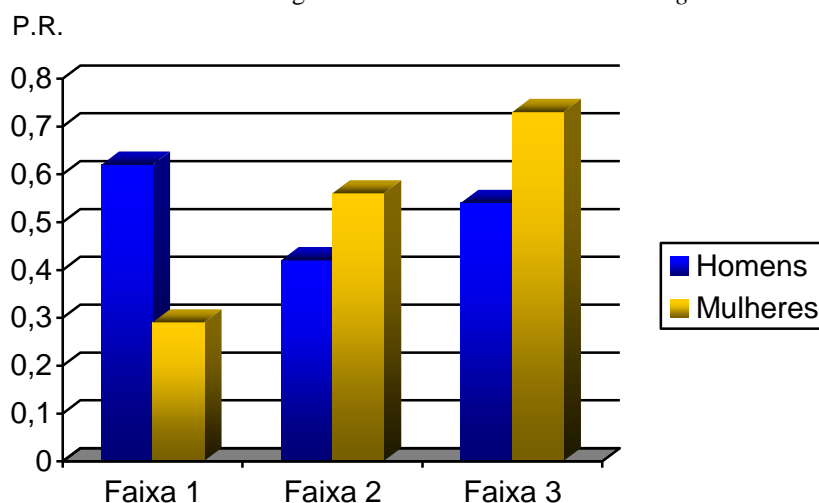
Tabela 9: Influência do gênero no uso de *a gente*.

Grupo: Gênero	Apl/T	%	P.R.
Homens	528/1369	39	0,45
Mulheres	621/1185	52	0,56
TOTAL	1149/2554	45	-

Nível de significância = 0,000

Os dados mostram que a tendência de as mulheres usarem a forma *a gente* é maior (0,56) do que a dos homens (0,45). Esse resultado confirma o que foi apontado em outros trabalhos, como o de Godoy (1999), o de Setti (1997), o de Menon (1994) e o de Cunha (1993), embora haja diferenças entre os *corpora* estudados.

É importante também analisar o grupo gênero conjugado com a faixa etária e com a localidade, para confirmarmos se são as mulheres da faixa 2 de Bananal / Barra dos Negros que estão fazendo um maior uso da forma *a gente*. Primeiramente, vejamos os resultados probabilísticos da conjugação do gênero com a faixa etária apresentados no gráfico 2.

Gráfico 3: Efeito do gênero e da faixa etária no uso de *a gente*.

Nível de significância = 0,002

O gráfico mostra que o uso da forma *a gente* varia ao considerarmos os fatores gênero e faixa etária conjuntamente. São as mulheres das faixas 2 seguidas das mulheres da faixa 1 que têm um uso mais significativo, com pesos relativos 0,66 e 0,60, respectivamente, o mesmo que acontece na comunidade de Bananal / Barra dos Negros, enquanto que os homens dessas faixas apresentam pesos relativos irrelevantes (0,43, na faixa 1, e 0,47, na faixa 2), o que já foi mostrado nos resultados das tabelas 8 e 9. Na faixa 3, há um certo equilíbrio no

comportamento lingüístico dos homens e das mulheres, porém com pesos relativos insignificantes: 0,39 e 0,38, respectivamente. Para melhor visualisarmos a atuação dessas variáveis, fizemos um cruzamento com o gênero, a localidade e a faixa etária. Os resultados encontram-se na tabela 10:

Tabela 10: Efeito do gênero, da faixa etária e da localidade no uso de *a gente*.

Gênero/ Localidade / Faixa etária		Bananal/Barra dos Negros			Piabas		
		Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	Faixa 1	123/296	42	0,49	56/180	31	0,41
	Faixa 2	144/341	42	0,42	86/206	42	0,41
	Faixa 3	107/276	39	0,42	12/70	17	0,24
Mulheres	Faixa 1	200/351	57	0,59	76/155	49	0,50
	Faixa 2	236/340	69	0,73	27/92	29	0,38
	Faixa 3	71/166	43	0,44	11/81	14	0,24

Nível de significância = 0,000

Essa tabela confirma que os maiores índices percentuais e probabilísticos de uso da forma *a gente* estão entre as mulheres da faixa 2 (69% e peso relativo 0,73), seguidas das da faixa 1 (57% e peso relativo 0,59) de Bananal / Barra dos Negros. Em Piabas, o uso de *a gente* entre os homens e entre as mulheres é insignificante.

Estes resultados confirmam que são as mulheres, principalmente da faixa 2, de Bananal / Barra dos Negros que utilizam com maior frequência a forma *a gente* para indeterminar o sujeito.

5.1.2 Nós

O pronome *nós* também é usado como uma das formas de se indeterminar o sujeito em português. Em nossos dados também podemos encontrá-lo. Na rodada em que colocamos esta forma como valor de aplicação, o programa excluiu também o grupo tipo de oração e mudança/manutenção do referente, como na rodada com a forma *a gente*. Além destes grupos, o programa também excluiu nessa rodada o grupo gênero do informante. Excluímos do arquivo de condições os grupos tipo de oração e manutenção/mudança do referente, visto que os resultados não foram sugestivos, e fizemos algumas modificações nos grupos restantes: 1) retiramos os tempos verbais presente e pretérito imperfeito do modo subjuntivo, por não apresentarem nenhuma ocorrência, e também o tempo futuro do presente do modo indicativo,

por apresentar apenas duas ocorrências; 2) amalgamamos todas as formas pronominais com sujeito explícito mais sujeito implícito, além de amalgamar as formas com sujeito vazio em um único fator, no grupo forma antecedente, como fizemos na rodada com a variante *a gente* como valor de aplicação; e 3) fizemos a combinação do grupo gênero com o grupo faixa etária.

Após fazer estas modificações, pudemos fazer uma segunda rodada e os grupos selecionados para o pronome *nós* foram:

- 1° Forma antecedente
- 2° Grau de indeterminação
- 3° Preenchimento do sujeito
- 4° Gênero e faixa etária
- 5° Localidade
- 6° Tempo e modo verbal
- 7° Tipo de verbo

5.1.2.1 Forma antecedente

Da mesma forma que ocorreu com o recurso *a gente*, a forma antecedente foi o primeiro grupo de fatores a ser selecionado pelo VARBRUL. Na tabela 11, vemos os resultados encontrados:

Tabela 11: Influência da forma antecedente no uso de *nós*.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	14/674	2	0,38
Nós	103/122	84	0,99
Você	1/107	1	0,41
Eles	1/160	1	0,11
Formas com sujeito zero	4/351	1	0,42
TOTAL	123/1291	9	-

Nível de significância = 0,000

Esta tabela mais uma vez confirma a hipótese de que em estruturas paralelas há uma tendência à manutenção da mesma forma. Conforme seus resultados, a probabilidade de o

falante usar *nós* é significativamente maior quando ele usa também *nós* numa oração que antecede esta forma (0,99).

Ao fazer um cruzamento com os grupos forma antecedente e mudança/ manutenção do referente³⁸ (este último mais uma vez não foi selecionado pelo programa estatístico), chegamos aos seguintes resultados:

Tabela 12: Efeito da forma antecedente e da mudança/ manutenção do referente no uso de *nós*.

Referência/ Forma antecedente	Igual		Diferente	
	Nº/T	%	Nº/T	%
A gente	11/634	2	3/40	8
Nós	102/112	91	0/10	0
Você	0/103	0	0/4	0
Eles	0/134	0	1/26	4
Formas com sujeito zero	1/302	0	3/49	6

Conforme a tabela 12, o uso da forma *nós* se dá quando o seu antecedente, numa seqüência discursiva, tem a mesma referência (91%). Se observarmos os números, a partir do momento que o falante muda de referente, há um favorecimento para o uso de uma forma distinta da precedente, porém são poucos os casos: apenas 7 ocorrências.

5.1.2.2 Grau de indeterminação

Assim como ocorreu com a forma *a gente*, o grau de indeterminação foi o segundo grupo selecionado pelo programa estatístico como favorecedor o uso de *nós*. Os resultados podem ser vistos na tabela 13:

³⁸ Fizemos esse cruzamento com todas as variantes da indeterminação estudadas neste trabalho e observamos sempre o mesmo resultado: a utilização da mesma forma em função da manutenção da referência e a mudança de forma em função da mudança de referência. Desta forma, achamos desnecessário continuar mostrando as tabelas, mostraremos apenas quando necessário.

Tabela 13: Influência do grau de indeterminação no uso de *nós*.

Grupo:			
Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	47/1132	4	0,34
Indeterminação parcial com referência implícita	149/1194	12	0,64
Indeterminação parcial com referência explícita	31/228	14	0,54
TOTAL	227/2554	9	-

Nível de significância = 0,000

O falante tende a utilizar *nós* quando a referência é parcial. A probabilidade é maior (0,64) quando a referência está implícita, mas é possível que seja recuperada a partir do contexto. Quando a referência está explícita no contexto, a probabilidade é menor (0,53), ficando próxima do ponto neutro (0,50). O uso de *nós* é pouco freqüente quando a indeterminação é completa (0,34). A partir dos dados dos *corpora*, observamos que o falante usa essa forma quando faz referência a ele mesmo e a outras pessoas, excluindo o interlocutor, reduzindo, assim, o seu conteúdo generalizador.

5.1.2.3 Preenchimento do sujeito

O grupo preenchimento do sujeito não foi muito significativo na amostra, sendo selecionado como favorável a aplicação da regra apenas com o indeterminador *nós*. Os resultados podem ser vistos na tabela 14:

Tabela 14: Influência do preenchimento do sujeito no uso de *nós*.

Grupo:			
Preenchimento do Sujeito	Apl/T	%	P.R.
Sujeito explícito	161/1173	14	0,54
Sujeito implícito	66/620	11	0,43
TOTAL	227/1793	13	-

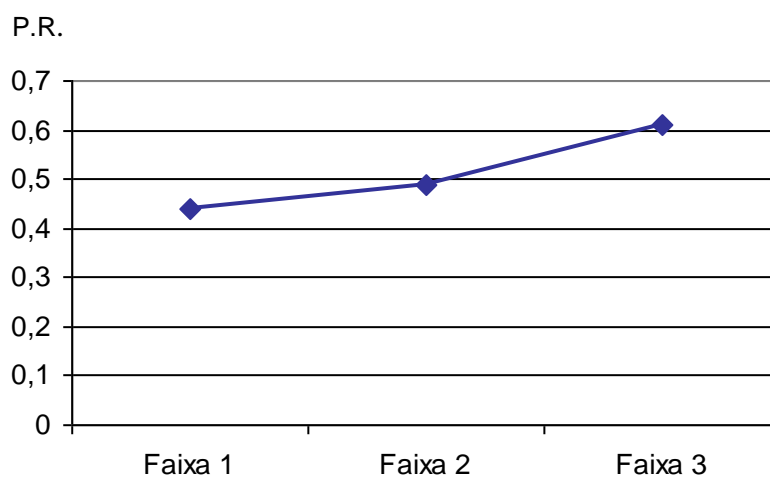
Nível de significância = 0,000

Vemos nesta tabela que o falante tem uma pequena tendência de usar a forma *nós* com sujeito explícito (0,54) em relação aos enunciados com sujeito implícito (0,43).

5.1.2.4 Gênero e faixa etária

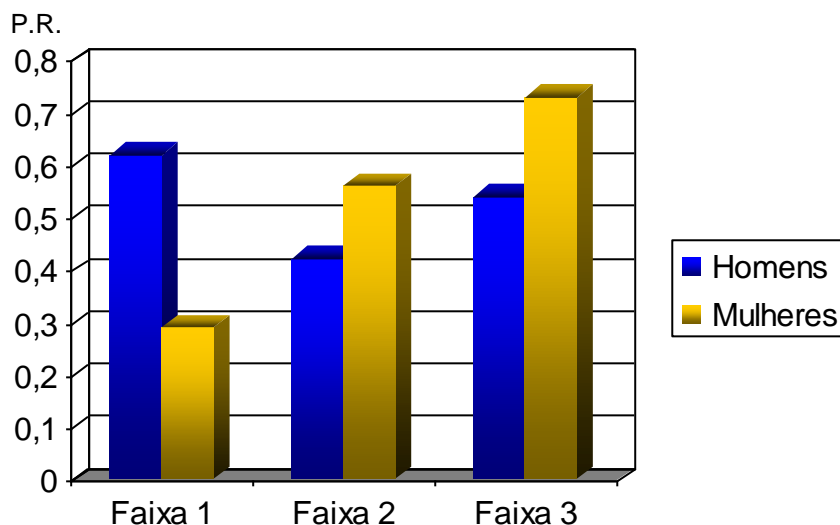
O gênero não foi selecionado pelo programa enquanto fator isolado na rodada geral, por isso, foi conveniente conjugá-lo com a faixa etária em um só grupo. Assim feito, o gênero associado a faixa etária foi selecionado pelo programa em quarto lugar. Antes de mostrarmos os resultados encontrados, apresentaremos o grupo faixa etária separadamente. Para isso, fizemos uma rodada com cada fator isoladamente e os resultados estão apresentados no gráfico 4.

Gráfico 4: Influência da faixa etária no uso de *nós*.



Nível de significância = 0,000

A partir dos resultados deste gráfico, percebemos um uso decrescente da forma *nós* na medida que a idade do informante diminui, sendo o uso mais significativo desta variante apenas na faixa 3 (0,61). Esses resultados mostram que o uso de *nós* nessas comunidades está passando por um processo de mudança em progresso, provavelmente em função do uso de *a gente*. Para sabermos se isso acontece de modo simétrico nas duas comunidades em estudo, faremos uma análise mais detalhada quando falarmos do fator localidade.

Gráfico 5: Efeito do gênero e da faixa etária no uso de *nós*.

Nível de significância = 0,000

Os resultados do gráfico 5 revelam que nas primeiras faixas etárias há um comportamento diferenciado entre homens e mulheres em relação ao uso da forma *nós*. Na faixa 1, os homens utilizam esta forma significativamente mais do que as mulheres (0,62 e 0,29, respectivamente). Nas faixas 2 e 3 ocorre uma inversão: as mulheres utilizam mais esta forma em relação aos homens.

Se observarmos os dados das mulheres, veremos que a frequência de uso da forma *nós* aumenta de modo regular em função da faixa etária das informantes, confirmando o uso cada vez menor desta variante pelas mulheres mais jovens. Em relação aos homens, o uso de *nós* é maior na faixa 3 (0,54), assim como entre as mulheres. Esses resultados mostram que tanto os homens quanto as mulheres da faixa 3 utilizam mais a forma *nós*. Se compararmos esses resultados com o gráfico 3, veremos que ocorre um processo inverso nessa faixa, pois homens e mulheres utilizam menos a forma *a gente*.

5.1.2.5 Localidade

Mais uma vez observamos um comportamento lingüístico diferenciado entre as duas localidades em análise.

Tabela 15: Influência da localidade no uso de *nós*.

Grupo: Localidade	Apl/T	%	P.R.
Bananal/Barra dos Negros	125/1770	7	0,44
Piabas	102/784	13	0,64
TOTAL	227/2554	9	-

Nível de significância = 0,000

Os resultados da tabela 15 indicam que Piabas favorece o uso da forma *nós* (0,64), diferente de Bananal / Barra dos Negros, onde o uso desta variante cai sensivelmente (0,44). Nesta última, como vimos na tabela 5, a preferência é para o uso de *a gente*.

É importante também conjugar a localidade com o gênero e com a faixa etária do informante a fim de observar mais especificamente esses resultados.

Tabela 16: Influência do gênero e da localidade no uso de *nós*.

Gênero/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	72/913	8	0,46	53/456	12	0,62
Mulheres	53/857	6	0,41	49/328	15	0,67
TOTAL	125/1770	7	-	102/784	13	-

Nível de significância = 0,007

Em relação ao gênero e à localidade, os resultados desta tabela confirmam que os pesos mais significativos para o uso de *nós* estão entre os falantes de Piabas, de forma mais expressiva entre as mulheres, que apresentam peso relativo 0,67. O uso desta forma entre os homens desta localidade é menor, 0,62. Em Bananal / Barra dos Negros, o uso de *nós* é insignificante entre homens e mulheres, seus pesos relativos estão abaixo de 0,50.

Em relação à faixa etária e à localidade, no uso desta forma, tivemos os seguintes resultados:

Tabela 17: Efeito da faixa etária e localidade no uso de *nós*.

Faixa Etária/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Faixa 1	41/647	6	0,42	28/335	8	0,54
Faixa 2	44/681	6	0,39	38/298	13	0,63
Faixa 3	40/442	9	0,55	36/151	24	0,79
TOTAL	125/1770	7	-	102/784	13	-

Nível de significância = 0,008

Vemos que o uso de *nós* em Piabas é significativo em todas as faixas etárias, havendo um uso decrescente (como vimos no gráfico 3) na medida que a idade do informante diminui. Em Bananal / Barra dos Negros o uso desta forma só é significativo na faixa 3 (0,55), o que é esperado nesta faixa, visto que são mais conservadores.

5.1.2.6 Tempo e modo verbal

Um outro fator selecionado como favorecedor para a aplicação de *nós* foi o tempo e modo verbal. Este, assim como com a forma *a gente*, foi o terceiro grupo a ser selecionado pelo VARBRUL e apresenta os resultados da tabela 18:

Tabela 18: Influência do tempo e modo verbal no uso de *nós*.

Grupo: Tempo e Modo Verbal	Apl/T	%	P.R.
Presente do indicativo	109/1630	7	0,44
Pretérito perfeito do indicativo	28/237	12	0,75
Pretérito imperfeito do indicativo	71/321	22	0,63
Futuro do subjuntivo	2/55	4	0,37
Infinitivo	13/274	5	0,51
Gerúndio	2/23	9	0,16
TOTAL	225/2540	9	-

Nível de significância = 0,000

A partir dos resultados desta tabela, vemos que, diferente do que acontece com o uso da forma *a gente*, o uso do indeterminador *nós* é maior com as formas de pretérito perfeito e imperfeito do indicativo, com pesos relativos 0,75 e 0,63, respectivamente, como também atestam Omena (1986), Lopes (1993) e Cunha (1993).

Fizemos uma análise cuidadosa dos contextos de uso da forma *nós* e vimos que esta predomina em trechos narrativos, em que os falantes versam sobre experiências vividas por eles, situando-se num momento anterior ao atual. Ao contar esses fatos, o informante se reporta a situações e tempos que pertencem a outras épocas, marcando cronologicamente esses fatos, o que caracteriza os tempos do pretérito.

5.1.2.7 Tipo de verbo

O tipo de verbo foi o grupo menos significativo no uso da variante *nós*, sendo o último selecionado pelo programa.

Tabela 19: Influência do tipo de verbo no uso de *nós*.

Grupo: Tipo de Verbo	Apl/T	%	P.R.
Transitivo	181/2155	8	0,50
Intransitivo	19/299	6	0,38
Cópula	27/100	27	0,79
TOTAL	227/2554	9	-

Nível de significância = 0,000

Como mostra a tabela 19, o uso de *nós* é mais importante quando o verbo é copulativo (0,79). Ao analisar esse grupo de fatores, vemos que ele não é tão importante na amostra e que está associado a outros fatores, como a saliência fônica da forma verbal. Dentre os 27 verbos copulativos usados com a forma *nós*, tem-se os verbos *ser*, *estar* e *ficar*, usados, em sua maioria, no presente do indicativo.

5.1.3 Você

O uso de *você* como indeterminador vem sendo comum na fala. De acordo com a análise de alguns trabalhos sobre indeterminação (cf. CUNHA, 1993; MENON, 1994; CAVALCANTE, 1998; GODOY, 1999), este é o indeterminador preferido pelos falantes cultos. Nos nossos *corpora*, a preferência de uso desta forma aparece em quinto lugar.

Ao fazer a rodada tendo a forma *você* como valor de aplicação, os grupos tipo de oração, tipo de verbo, preenchimento do sujeito, mudança/ manutenção do referente e localidade não foram selecionados pelo programa. Sendo assim, tivemos que excluir estes grupos da rodada e fazer algumas modificações: 1) excluímos os tempos e modos verbais futuro do presente do indicativo, presente e pretérito imperfeito do subjuntivo e a forma nominal de gerúndio, devido à ausência ou pequeno número de ocorrências; 2) amalgamamos as variantes indeterminação parcial com referência implícita no contexto com indeterminação parcial com referência explícita no contexto numa única variante e damos o nome de indeterminação parcial; e 3) amalgamamos todas as formas pronominais de sujeito explícito

com sujeito implícito, além de amalgamarmos as formas com sujeito vazio em um único fator, no grupo forma antecedente.

Assim feito, fizemos a rodada e tivemos os seguintes grupos selecionados:

- 1° Forma antecedente
- 2° Grau de indeterminação
- 3° Faixa etária
- 4° Tempo e modo verbal
- 5° Gênero

5.1.3.1 Forma antecedente

Mais uma vez o grupo forma antecedente é estatisticamente significativo, sendo o primeiro grupo selecionado. Os resultados da tabela 20 indicam que a presença de *você* como forma antecedente numa seqüência discursiva tende a favorecer o uso da mesma forma em orações posteriores (0,99).

Tabela 20: Influência da forma antecedente no uso de *você*.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	3/674	0	0,31
Nós	1/122	1	0,46
Você	98/107	92	0,99
Formas com sujeito zero	3/351	1	0,41
TOTAL	105/1254	8	-

Nível de significância = 0,000

5.1.3.2 Grau de indeterminação

O grau de indeterminação foi o segundo grupo selecionado. Os resultados estatísticos estão expostos na tabela 21:

Tabela 21: Influência do grau de indeterminação no uso de *você*.

Grupo: Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	140/1132	12	0,81
Indeterminação parcial	12/1422	1	0,24
TOTAL	152/2554	6	-

Nível de significância = 0,000

Observamos nesta tabela que a forma pronominal *você* apresenta maior probabilidade quando o grau de indeterminação é completo (0,81), indicando que o falante usa essa forma quando há uma “extensão” de pessoas envolvidas nesse tipo de referência, a qual denominamos de genérica, que inclui todas as pessoas do discurso. Nos nossos dados, essa abrangência tem um sentido mais amplo ou mais restrito, a depender do seu referente. Vejamos os exemplos (88) e (89):

(88) Doc.: E a farinha, como é que faz?

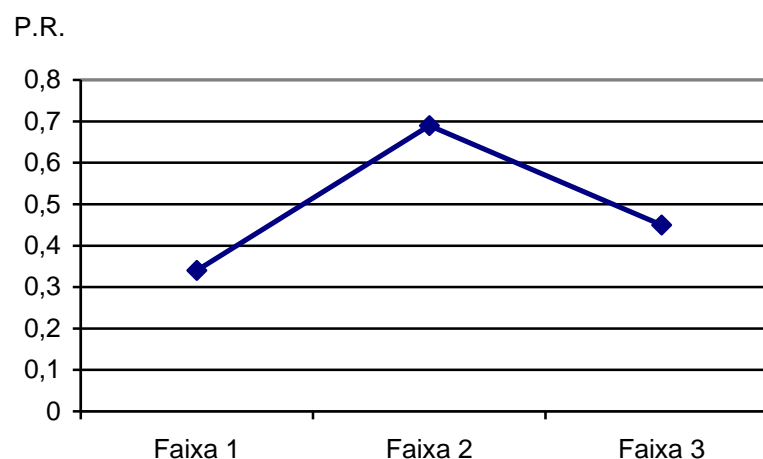
Inf.: A farinha também é aquele mermo tipo da... da rapadura, né? **Você** tem... quem tem o motor, hoje em dia é mai fácil, né? Aí **você** passa na... a mandioca ali no motor, aí aquela massa já sai, já fica lá na frente, né? Ali depois **cê** pega aquela massa, tem uma prensa. Aí **cê** vai, panha aquela massa na prensa, aí **Ø** enxuga ela, ai depois de tudo, **cê** vem, quem tem motor vem, quem num tem então... num... numa peneira assim, né? Aí **cê** vai, passa ela na peneira, e aí **Ø** põe no forno. É a merma coisa da... da farinha. (B, 9, 143-4).

(89) Inf.: No tempo do cru... do conto, **você** com um conto era a merma coisa que num ter nada, **cê** ia na feira com aquele conto, **cê** num comprava nada... (B, 9, 133).

No exemplo (88), a referência é mais ampla e entendida como a soma de todas as pessoas do discurso. Quando o informante explica como se faz farinha, ele fala por ele, pelo seu locutor e por todo mundo, sem restrição a um grupo social ou a um período específico. Já no exemplo (89), a forma *você* usada pelo informante, apesar de envolver também todas as pessoas do discurso, tem sua abrangência reduzida, pois o informante restringe às pessoas que tivessem vivido em uma época anterior àquela dos interlocutores do diálogo.

5.1.3.3 Faixa etária

A faixa etária foi o terceiro grupo selecionado pelo VARBRUL como favorecedor a aplicação da forma *você*.

Gráfico 6: Influência da faixa etária no uso de *você*.

Nível de significância = 0,000

O gráfico 6 mostra que o uso de *você* como indeterminador ocorre com maior frequência na faixa 2 (0,69), sendo a faixa 1 a que menos usa o pronome (0,34). Esses resultados não correspondem aos resultados de outros trabalhos sobre este tema que analisaram a forma *você* (cf. CUNHA, 1993; SETTI, 1997; GODOY, 1999). Nestes, apesar de analisar outros *corpora*, esta forma é mais usada pelos falantes da faixa 1, justamente por ser uma forma inovadora para indeterminar o sujeito.

Quando conjugamos este grupo com a localidade, vimos que o seu uso é intensamente maior entre os falantes da faixa 2 de Bananal / Barra dos Negros, com um peso relativo 0,80, apresentando o maior número de ocorrências: 106 das 152 ocorrências de *você* nos *corpora* (correspondendo a 70% deste total).

Tabela 22: Efeito da faixa etária e localidade no uso de *você*.

Faixa Etária/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Faixa 1	9/647	1	0,32	7/335	2	0,42
Faixa 2	106/681	16	0,80	2/298	1	0,25
Faixa 3	24/442	5	0,45	4/151	3	0,60
TOTAL	139/1770	9	-	13/784	2	-

Nível de Significância = 0,002

Fizemos uma investigação nos dados e verificamos que o informante 9 de Bananal / Barra dos Negros, gênero masculino, faixa etária 2, produziu esta variante 96 vezes durante a entrevista. Se considerarmos que são 152 ocorrências desta forma em 24 entrevistas, o que dá

uma frequência média de 7 por informante, podemos dizer que um único informante exerceu uma influência muito acentuada no resultado de *você* na fala dos informantes da faixa 2 de Bananal / Barra dos Negros. Em Piabas, o uso de *você* só é significativo na faixa 3, porém só há o registro de 4 ocorrências desta forma entre os falantes desta comunidade.

5.1.3.4 Tempo e modo verbal

Em relação a esta variável, observamos, conforme os dados da tabela 23, que a variante *você* é mais usada ao lado de verbos no futuro do subjuntivo e presente do indicativo, com pesos relativos 0,76 e 0,55, respectivamente.

Tabela 23: Influência do tempo e modo verbal no uso de *você*.

Grupo:			
Tempo e Modo Verbal	Ap/T	%	P.R.
Presente do indicativo	121/1630	7	0,55
Pretérito perfeito do indicativo	2/237	1	0,08
Pretérito imperfeito do indicativo	14/321	4	0,41
Futuro do subjuntivo	3/55	5	0,76
Infinitivo	10/274	4	0,38
TOTAL	150/2517	6	-

Nível de significância = 0,008

O futuro do subjuntivo é o tempo mais usado com a forma *você*, mas se analisarmos o número de ocorrências, veremos que é muito pequeno em relação ao presente do indicativo, são apenas 3 para aquele e 121 para este, e também à outros tempos não significativos para o programa. Observamos nos *corpora* os contextos em que o presente do indicativo ocorre e verificamos que, assim como foi mostrado no uso da forma *a gente*, este tempo é usado pelos informantes sempre que eles querem descrever ou explicar ações habituais ou frequentes, não havendo marcação temporal, o que propicia o uso de *você*. Vejamos um exemplo em que esse tempo ocorre com a forma *você*:

(90) Doc.: E qual a roça mais fácil de trabalhar? (...)

Inf.: Feijão é...

Doc.: Hum.

Inf.: ... **você pranta** ele, com noventa dia **você tá colheno** ele no pé. (B, 13, 173)

5.1.3.5 Gênero

Os resultados da tabela 24 indicam que os homens utilizam significativamente mais a forma *você* do que as mulheres (0,59 e 0,40, respectivamente).

Tabela 24: Influência do gênero no uso de *você*.

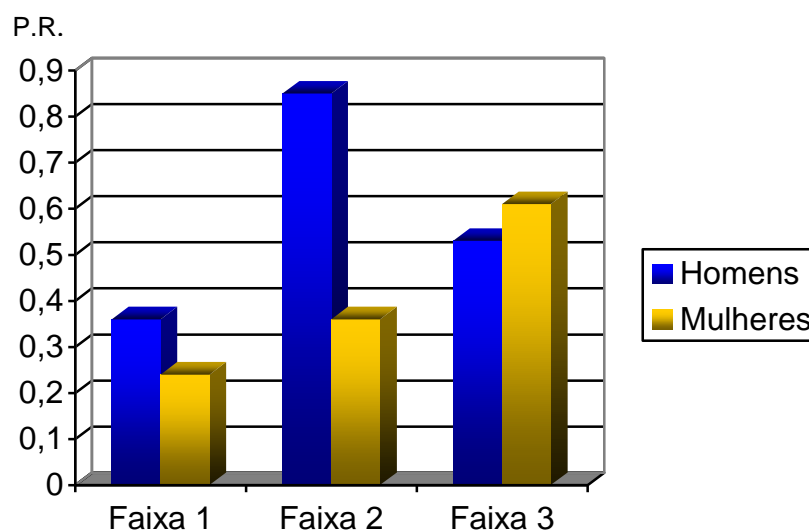
Grupo: Gênero	Ap/T	%	P.R.
Homens	123/1369	9	0,59
Mulheres	29/1185	2	0,40
TOTAL	152/2554	6	-

Nível de significância = 0,009

Porém, é necessário lembrar que o grande número de ocorrências entre os homens, 123 contra 29 entre as mulheres, deve-se ao fato de um único informante do gênero masculino de Bananal / Barra dos Negros utilizar esta forma 96 vezes, restando apenas 27 ocorrências para os outros três informantes do gênero masculino que compõem a amostra.

A fim de observarmos os resultados destas variáveis em conjunto, resolvemos fazer uma rodada conjugando-as. Os pesos relativos constam no gráfico 7.

Gráfico 7: Efeito do gênero e da faixa etária no uso de *você*.



Nível de significância = 0,000

Se observarmos este gráfico, veremos que são os homens da faixa 2 que mais utilizam a forma *você*, justamente devido a um informante que usa intensamente esta forma. Já na faixa 3, a frequência é equilibrada, com uma tendência de uso maior por parte das mulheres. Em

relação a estas, elas empregam mais o pronome conforme a faixa etária avança. Sendo assim, apresentam um aumento gradativo e constante.

Tabela 25: Efeito do gênero e localidade no uso de *você*.

Gênero/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	119/913	13	0,72	4/456	4	0,25
Mulheres	20/857	2	0,37	9/328	9	0,57
TOTAL	139/1770	8	-	13/784	2	-

Nível de significância = 0,000

A tabela 25 confirma mais uma vez a influência do informante do gênero masculino de Bananal / Barra dos Negros no uso de *você*, com o maior índice percentual e peso relativo (13% e 0,72, respectivamente). Em Piabas, o uso é maior entre as mulheres (0,57). Com exceção do citado informante masculino de Bananal / Barra dos Negros, o uso desta variante pelos informantes que compõem os *corpora* é muito pequeno. Sendo assim, não podemos fazer uma afirmação em relação ao uso de *você* nestes *corpora* com 2/3 do total de ocorrências usadas por apenas um entrevistado.

5.1.4 Eles

Em nossos *corpora*, o pronome *eles* é também usado como uma forma para se indeterminar o sujeito. Na rodada que tivemos como valor de aplicação esta variante, o programa excluiu como não favorecedores a aplicação da regra os grupos: tipo de oração, tipo de verbo, preenchimento do sujeito, mudança/manutenção do referente e localidade. Esses grupos foram excluídos do arquivo de condição e algumas modificações nos grupos de fatores restantes foram feitas: 1) os tempos verbais presente e pretérito imperfeito do subjuntivo foram excluídos devido à ausência de ocorrências para a variante em análise; 2) amalgamamos todas as formas pronominais com sujeito explícito com as de sujeito implícito, além de excluir o fator $\emptyset+V+SE$, por não apresentar dados no grupo forma antecedente; 3) amalgamamos gênero e faixa etária num único grupo.

Feitas as modificações, os grupos selecionados como favorecedores a aplicação da forma *eles* foram apenas quatro:

1º Forma antecedente

2º Grau de indeterminação

3° Tempo e modo verbal

4° Gênero e faixa etária

5.1.4.1 Forma antecedente

Da mesma forma que ocorre nas rodadas com outras formas pronominais, ocorre com o indeterminador *eles*: a forma antecedente é o primeiro grupo selecionado estatisticamente como favorecedor a aplicação da regra. Os resultados constam na tabela 26:

Tabela 26: Influência da forma antecedente no uso de *eles*.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	16/674	2	0,39
Nós	1/122	1	0,07
Você	1/107	1	0,28
Eles	133/160	83	0,99
Formas com sujeito vazio	12/351	3	0,39
TOTAL	163/1414	12	-

Nível de significância = 0,000

A partir desses resultados, podemos observar que mais uma vez a hipótese da mesma forma numa seqüência discursiva é comprovada. A forma *eles* é a mais usada quando o seu antecedente também é *eles* (0,99). Apesar do baixo valor probabilístico (0,39), ao usar esta forma para indeterminar o sujeito, o falante usa as formas com sujeito vazio e também a forma *a gente* como forma antecedente numa freqüência maior do que com as outras variantes da indeterminação em que este grupo foi selecionado.

5.1.4.2 Grau de indeterminação

Este foi o segundo grupo mais importante considerado pelo programa para a aplicação da forma em análise. Os resultados estatísticos podem ser vistos na tabela 27.

Tabela 27: Influência do grau de indeterminação no uso de *eles*.

Grupo:			
Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	68/1132	6	0,53
Indeterminação parcial com referência implícita no contexto	91/1194	8	0,40
Indeterminação parcial com referência explícita no contexto	113/228	50	0,81
TOTAL	272/2569	11	-

Nível de significância = 0,005

Os números desta tabela indicam que a indeterminação parcial com referência explícita no contexto é a que mais favorece o uso de *eles* como forma para indeterminar o sujeito. O peso relativo, 0,81, diferencia-se significativamente dos outros fatores. Vimos no item 4.4.1.4 que esta forma é exclusiva de terceira pessoa, excluindo a primeira e a segunda, além de sempre fazer referência a um grupo socialmente bem delimitado.

A alta probabilidade de uso dessa forma na indeterminação parcial com referência explícita no contexto se dá como no exemplo (91):

(91) Doc.: E mataro por quê?

Inf.: De serem miserave, se porque ele tinha... ele tra... é o nome era E. L. Agora trata o Q. Porque diz que Q. tinha dedado a poliça que **eles** era bandido. (P, 12, 182)

Nesse exemplo, a referência de *eles* envolve um grupo específico (os bandidos) que não inclui os interlocutores da ação. Assim, a generalização obtida não é total, sua abrangência é mais reduzida do que aquelas apresentadas por outras formas pronominais usadas na indeterminação.

5.1.4.3 Tempo e modo verbal

Os resultados do terceiro grupo selecionado constam na tabela 28.

Tabela 28: Influência do tempo e modo verbal no uso de *eles*.

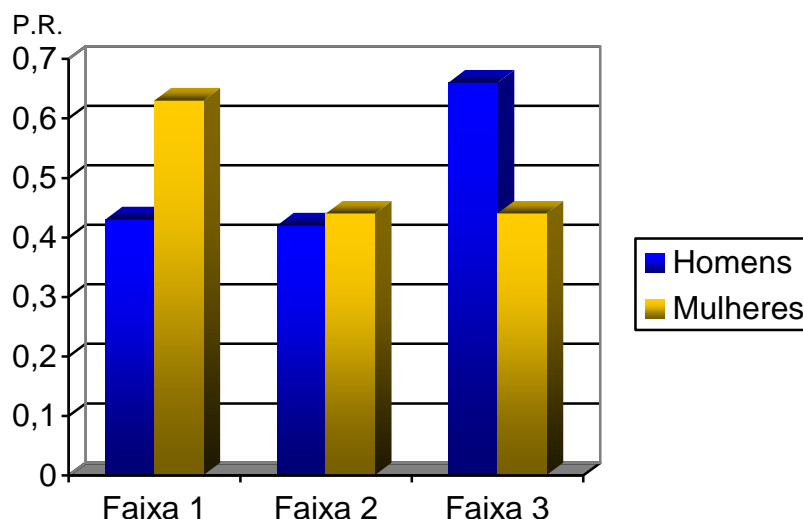
Grupo: Tempo e Modo Verbal	Apl/T	%	P.R.
Presente do indicativo	183/1630	11	0,59
Pretérito perfeito do indicativo	32/237	14	0,43
Pretérito imperfeito do indicativo	39/321	12	0,38
Futuro do subjuntivo	1/55	2	0,17
Infinitivo	15/274	5	0,28
Gerúndio	2/23	9	0,30
TOTAL	272/2540	11	-

Nível de significância = 0,000

Observamos que o tempo verbal que mais favorece a forma *eles* é o presente do indicativo (0,59). Assim como evidenciamos no uso das formas *a gente* e *você*, verificamos as ocorrências de *eles* e vimos que o uso desta forma com o tempo presente do indicativo pode ser justificado exatamente pela tipologia do discurso. Os contextos argumentativos e descritivos são os mais utilizados com esta forma. Nestes, não há delimitação de duração, fato que caracteriza o presente do indicativo.

5.1.4.4 Gênero e faixa etária

O gênero analisado separadamente não atuou de modo significativo, sendo excluído pelo programa. Procedemos à combinação deste com a faixa etária que juntos constituíram um novo grupo de fatores.

Gráfico 8: Efeito do gênero e da faixa etária no uso de *eles*.

Nível de significância = 0,005

Como mostra o gráfico 8, há um uso invertido desta forma entre homens e mulheres das faixas 1 e 3. Na faixa 1, são as mulheres que mais utilizam esta forma, com peso relativo 0,63; já os homens têm um uso significativo na faixa 3, com peso relativo 0,66. Na faixa 2, se observarmos o peso relativo, há um uso equilibrado entre os dois gêneros: 0,42 entre os homens e 0,44 entre as mulheres. Quando analisamos isoladamente esses fatores, percebemos um uso um pouco maior da variante *eles* por parte das mulheres (13%) em relação aos homens (9%).

Fizemos um cruzamento entre essas variáveis e a variável localidade. Esta, quando combinada com o gênero, foi excluída pelo programa, pois não há diferença no uso desta variante entre as comunidades: homens e mulheres apresentam o mesmo percentual de uso. Já com a faixa etária, a rodada apresentou os resultados da tabela 29:

Tabela 29: Efeito da faixa etária e da localidade no uso de *eles*.

Faixa Etária/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Faixa 1	99/647	15	0,55	41/335	12	0,51
Faixa 2	39/681	6	0,34	32/298	11	0,60
Faixa 3	48/442	11	0,63	13/151	9	0,44
TOTAL	186/1770	10	-	86/784	11	-

Nível de significância = 0,002

Se observarmos os percentuais desta tabela, veremos que, apesar de baixos, os falantes das faixas 1 e 3 de Bananal / Barra dos Negros são os que mais utilizam a forma *eles*, com uma frequência maior na faixa 1. Em Piabas, o seu uso vai aumentando na medida que a faixa etária diminui. Os pesos relativos mostram que as faixas 1 e 3 de Bananal / Barra dos Negros têm a maior probabilidade de uso desta forma, já em Piabas, o uso desta variante é mais significativo na faixa 2.

5.1.5 Ø+V3PS

A forma de indeterminação Ø+V3PS é a segunda forma preferida pelos falantes das comunidades. Nas rodadas estatísticas, vimos que tanto os fatores lingüísticos quanto os fatores sociais condicionam o seu uso.

Alguns grupos de fatores, como tipo de oração, tipo de verbo, preenchimento do sujeito, mudança/manutenção do referente e localidade, foram considerados irrelevantes a

aplicação da forma $\emptyset+V3PS$, sendo eliminados pelo programa. Nos grupos selecionados, algumas alterações foram necessárias, como: 1) excluímos os tempos presente e pretérito imperfeito do modo subjuntivo e futuro do presente do indicativo devido à ausência ou pequena quantidade de ocorrências; 2) amalgamamos as variantes indeterminação parcial com referência implícita no contexto com indeterminação parcial com referência explícita no contexto numa única variante, assim como fizemos com a forma *você*; 3) amalgamamos todas as formas pronominais de sujeito explícito e sujeito implícito, além de excluir o fator $\emptyset+V+SE$, pois não apresentou dados no grupo forma antecedente.

Após estas modificações, os grupos considerados estatisticamente relevantes pelo programa foram:

- 1° Forma antecedente
- 2° Tempo e modo verbal
- 3° Grau de indeterminação
- 4° Gênero
- 5° Faixa etária

5.1.5.1 Forma antecedente

O primeiro grupo selecionado como favorecedor para o uso da forma $\emptyset+V3PS$ foi a forma antecedente. Na tabela 30, apresentamos os resultados:

Tabela 30: Influência da forma antecedente no uso do $\emptyset+V3PS$.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	13/674	2	0,23
Nós	2/122	2	0,22
Você	2/107	2	0,15
Eles	5/160	3	0,37
$\emptyset+V3PS$	198/246	80	0,99
$\emptyset+V3PP$	3/51	6	0,45
$\emptyset+VINF$	17/52	33	0,95
TOTAL	240/1172	17	-

Nível de significância = 0,000

A partir dos resultados desta tabela, mais uma vez é confirmado o paralelismo entre as formas. O peso relativo é significativo quando a forma $\emptyset+V3PS$ é antecedida por $\emptyset+V3PS$,

com peso relativo 0,99. A forma $\emptyset+VINF$ quanto antecede a forma $\emptyset+V3PS$ tem também um uso significativo (0,95). Acreditamos que seu alto índice probabilístico ocorre devido a intercambialidade dessas formas num mesmo contexto, ambas têm o mesmo grau de abrangência: englobam as três pessoas do discurso.

Nos nossos dados, são comuns exemplos do tipo:

(92) Inf.: Eu digo, pra \emptyset **sair** daqui, por exemplo, pra \emptyset **ir** pro Calundu \emptyset **vai** pelo Gonçalo aí, né?

Doc.: Hum...

Inf.: Mas pra quem já tá lá, passa por lá mesmo. (P, 7, 142).

5.1.5.2 Tempo e modo verbal

O Tempo e modo verbal foi o segundo grupo de fatores mais significativo para o uso da variante $\emptyset+V3PS$. Os resultados numéricos constam na tabela 31:

Tabela 31: Influência do tempo e modo verbal no uso do $\emptyset+V3PS$.

Grupo:			
Tempo e Modo Verbal	Apl/T	%	P.R.
Presente do indicativo	374/1630	23	0,59
Pretérito perfeito do indicativo	40/238	17	0,42
Pretérito imperfeito do indicativo	65/321	20	0,61
Futuro do subjuntivo	27/55	49	0,85
Infinitivo	14/274	5	0,08
Gerúndio	1/23	4	0,07
TOTAL	521/2540	21	-

Nível de significância = 0,000

A partir dos resultados desta tabela, vemos que a forma $\emptyset+V3PS$ é usada preferencialmente com o futuro do subjuntivo (0,85), pretérito imperfeito do indicativo (0,61) e presente do indicativo (0,59).

Novamente, ao analisar os contextos em que esta forma é usada, a maioria das ocorrências está em enunciados que indicam condição, referência ao passado e, em sua grande parte, descrição e argumentação de ações habituais. Estes, como já vimos na análise de outras variantes da indeterminação, são contextos que propiciam o uso dos tempos verbais mais significativos para ocorrência da forma $\emptyset+V3PS$.

5.1.5.3 Grau de indeterminação

O terceiro grupo selecionado pelo programa como favorecedor para o uso da forma $\emptyset+V3PS$, foi o grau de indeterminação. Essa variante apresenta maior frequência e peso relativo quando a indeterminação é completa (30% e 0,65, respectivamente), conforme vemos na tabela 32:

Tabela 32: Influência do grau de indeterminação no uso do $\emptyset+V3PS$.

Grupo: Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	336/1132	30	0,65
Indeterminação parcial	189/1422	13	0,38
TOTAL	525/2554	21	-

Nível de significância = 0,000

Esta forma é usada quando o falante faz referência às três pessoas do discurso. Em nossos dados, a maioria das ocorrências aparece como em (93):

(93) Doc.: Como é que faz aqui tijolo?

Inf.: Ali é... \emptyset cavaca a terra, \emptyset móia, ma... \emptyset móia bem moiadim e \emptyset massa o barro bem massadinho mó dele ficar assim liguento. Aí, \emptyset põe na forma e \emptyset soca tudo, aí forma o adobo. (B, 7, 80).

Nesse exemplo, o informante usa de modo genérico a forma $\emptyset+V3PS$. Ao explicar como se faz tijolo, não se restringe a ele, e sim ao modo como qualquer pessoa poderia fazê-lo.

5.1.5.4 Gênero

O gênero do informante foi o quarto grupo selecionado como favorecedor para o uso da variante em análise. A tabela 33 mostra que os homens usam relativamente mais a forma $\emptyset+V3PS$ do que as mulheres (pesos relativos 0,56 e 0,43, respectivamente).

Tabela 33: Influência do gênero no uso do $\emptyset+V3PS$.

Grupo: Gênero	Apl/T	%	P.R.
Homens	336/1369	25	0,56
Mulheres	189/1185	16	0,43
TOTAL	525/2554	21	-

Nível de significância = 0,000

Conjugamos esse grupo de fatores com o grupo localidade (que não foi selecionado pelo programa para essa variante) para averiguar se esse comportamento lingüístico se mantém em cada localidade considerada isoladamente e obtivemos os resultados que constam na tabela 34:

Tabela 34: Efeito do gênero e da localidade no uso do $\emptyset+V3PS$.

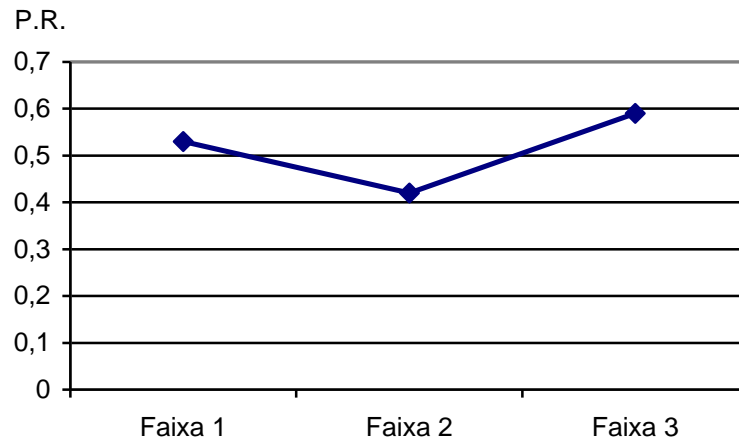
Gênero/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	212/913	23	0,55	124/456	27	0,59
Mulheres	126/857	15	0,43	63/328	19	0,42
TOTAL	338/1770	19	-	187/784	24	-

Nível de significância = 0,001

Nesta tabela, vemos que tanto os homens de Bananal / Barra dos Negros quanto os de Piabas utilizam de forma significativa a forma $\emptyset+V3PS$, com uma probabilidade maior em Piabas (0,59 e 0,55 em Bananal / Barra dos Negros). Entre as mulheres, o uso desta forma é insignificante nas duas comunidades. Observamos, assim, que não há diferenças lingüísticas entre as comunidades no uso que seus falantes fazem da variante $\emptyset+V3PS$.

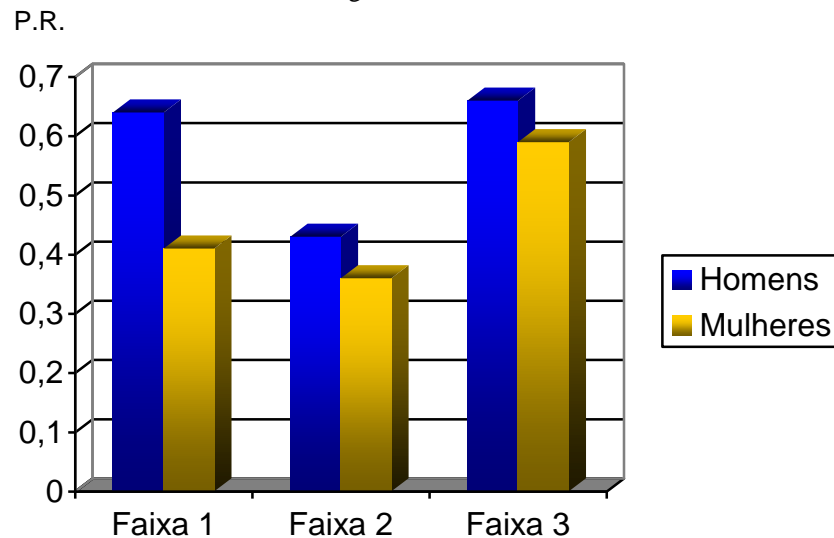
5.1.5.5 Faixa etária

Em relação à faixa etária, a faixa 3 é a que mais favorece o uso de $\emptyset+V3PS$, tanto em percentual quanto em peso relativo (29% e 0,59, respectivamente), seguida da faixa 1 (22% e 0,53). Na faixa 2, o uso desta forma é insignificante (0,42). Vejamos os pesos relativos no gráfico 9:

Gráfico 9: Influência da faixa etária no uso do $\emptyset+V3PS$.

Nível de significância = 0,000

Quando cruzamos esse grupo com o gênero do informante, obtivemos os resultados do gráfico 10:

Gráfico 10: Efeito do gênero e da faixa etária no uso de $\emptyset+V3PS$.

Nível de significância = 0,004

Esse gráfico confirma os resultados vistos nas tabelas 29 e 31: os homens utilizam mais a forma $\emptyset+V3PS$ do que as mulheres em todas as faixas etárias, principalmente nas faixas 3 (0,66) e 1 (0,60).

5.1.6 Ø+V3PP

A forma indeterminadora Ø+V3PP é consagrada pela GT como uma das formas utilizadas para indeterminar o sujeito em português. Em nossos dados, também encontramos ocorrências desta forma, porém numa frequência menor se compararmos com o uso de outras formas que a GT não menciona. Nas rodadas estatísticas, vimos que tantos os fatores lingüísticos quanto os fatores sociais condicionam o seu uso.

A maioria dos grupos de fatores não foi selecionada pelo programa, como tipo de oração, tipo de verbo, preenchimento do sujeito, mudança/manutenção do referente e os fatores sociais gênero e faixa etária, e nos grupos selecionados tivemos que fazer algumas alterações, como: 1) exclusão dos tempos presente e pretérito perfeito do modo subjuntivo, futuro do presente do modo indicativo e as formas nominais gerúndio e infinitivo, devido à ausência de ocorrências; 2) exclusão das formas pronominais *você* e *nós* tanto com sujeito explícito quanto com implícito e do fator Ø+V+SE, devido à ausência de dados.

Apenas quatro grupos foram selecionados como relevantes. Na ordem, temos:

- 1º Tempo e modo verbal
- 2º Forma antecedente
- 3º Grau de indeterminação
- 4º Localidade

5.1.6.1 Tempo e modo verbal

Na rodada com a forma Ø+V3PP como aplicação da regra variável, o tempo e modo verbal foi o primeiro grupo de fatores selecionado. A partir dos resultados da tabela 35, podemos constatar que esta forma é usada preferencialmente com o pretérito perfeito do indicativo (0,96).

Tabela 35: Influência do tempo e modo verbal no uso do Ø+V3PP.

Grupo:			
Tempo e Modo Verbal	Apl/T	%	P.R.
Presente do indicativo	26/1630	2	0,44
Pretérito perfeito do indicativo	90/237	38	0,96
Pretérito imperfeito do indicativo	2/321	1	0,28
Futuro do subjuntivo	2/55	4	0,43
TOTAL	120/2243	5	-

Nível de significância = 0,000

Isso se dá porque a forma $\emptyset+V3PP$ ocorre em contextos reais em que há a desfocalização do sujeito e o que se evidencia é o conteúdo expresso pelo verbo. Assim, em

(94) Doc.: Hum... hum... negócio de ladrão, essas coisa, tem por tem por aqui?

Inf.: Num senhora, eu nunca vi... só vejo o povo falano... o povo fala que diz que... agora \emptyset **sortaro** um bocado, mai graças a Deus que nunca \emptyset **prendero** pó lado daqui não. (P, 8, 121)

temos a ação pura e simples e o fato veiculado por ‘sortaro’ e ‘prendero’ é considerado como um todo único, analisável globalmente.

5.1.6.2 Forma antecedente

A forma antecedente mostrou-se relevante no uso desta variante e, conforme podemos ver na tabela 36, mais uma vez é confirmado o paralelismo existente no uso das formas de indeterminação. O peso relativo é significativamente maior quando a forma que antecede $\emptyset+V3PP$ é também $\emptyset+V3PP$, com alto índice percentual e probabilístico (78% e 0,99, respectivamente).

Tabela 36: Influência da forma antecedente no uso do $\emptyset+V3PP$.

Grupo: Forma Antecedente	Apl/T	%	P.R.
A gente	3/674	0	0,48
Eles	1/160	1	0,36
$\emptyset+V3PS$	4/246	2	0,46
$\emptyset+V3PP$	40/51	78	0,99
TOTAL	48/1131	4	-

Nível de significância = 0,000

5.1.6.3 Grau de indeterminação

O grau de indeterminação foi o terceiro grupo selecionado pelo programa para o uso do $\emptyset+V3PP$. Os resultados constam na tabela 37:

Tabela 37: Influência do grau de indeterminação no uso do $\emptyset+V3PP$.

Grupo: Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	91/1132	8	0,71
Indeterminação parcial com referência implícita	19/1194	2	0,32
Indeterminação parcial com referência explícita	10/228	4	0,39
TOTAL	120/2554	5	-

Nível de significância = 0,000

Conforme os resultados desta tabela, a indeterminação completa é o fator que mais favorece o uso de $\emptyset+V3PP$, com peso relativo 0,71. Apesar de ser uma forma que envolve apenas a ‘não-pessoa’, ela tem uma abrangência genérica e exprime a imprecisão de idéia ou juízo que fazemos de qualquer pessoa.

Abaixo, temos um exemplo retirado dos nossos *corpora*:

(95) Doc.: Pau de rato é o que?

Inf.: O pau de rato é um... um mato que tem no mato Cuma diz...

Doc.: Hum...

Inf.: É remédio.

Doc.: Serve pra que ele?

Inf.: \emptyset **Dize** que é bom pa de comer que faz mal. (P, 8, 114)

Neste exemplo, não é possível identificar quem realizou a ação expressa pela forma ‘dize’, remetendo a um grande número de pessoas indeterminadas.

5.1.6.4 Localidade

O último grupo selecionado foi a localidade. A tabela 38 mostra os números encontrados:

Tabela 38: Influência da localidade no uso do $\emptyset+V3PP$.

Grupo: Localidade	Apl/T	%	P.R.
Bananal/Barra dos Negros	37/1170	2	0,40
Piabas	83/794	11	0,70
TOTAL	120/2554	5	-

Nível de significância = 0,000

O resultado referente a esse grupo de fatores indica que os falantes da comunidade de Piabas utilizam com maior probabilidade a forma $\emptyset+V3PP$ do que os falantes de Bananal / Barra dos Negros (0,70 e 0,40, respectivamente).

Fizemos um cruzamento entre o gênero do informante e a localidade para verificarmos como se comportam lingüisticamente os homens e as mulheres das comunidades em estudo, mas este foi excluído pelo programa VARB2000, pois os resultados foram irrelevantes. Na tabela 39, expomos esses resultados a partir dos pesos relativos obtidos no nível 1³⁹ desta rodada:

Tabela 39: Influência do gênero e da localidade no uso do $\emptyset+V3PP$.

Gênero/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	21/913	2	0,41	44/456	10	0,65
Mulheres	16/857	2	0,42	39/328	12	0,70
TOTAL	37/1770	2	-	83/784	10	-

Esta tabela comprova que tanto os homens quanto as mulheres das duas comunidades têm um uso equilibrado dessa variante, porém é em Piabas que há um uso significativo, comprovando os resultados da tabela 38.

Cruzamos também as variáveis localidade e faixa etária do informante e tivemos os seguintes resultados:

Tabela 40: Efeito da faixa etária e da localidade no uso do $\emptyset+V3PP$.

Faixa Etária/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Faixa 1	18/647	3	0,55	20/335	6	0,68
Faixa 2	12/681	2	0,34	31/298	10	0,66
Faixa 3	7/442	2	0,31	32/151	21	0,80
TOTAL	37/1770	2	-	83/784	10	-

Nível de significância = 0,000

Em Bananal / Barra dos Negros, a maior probabilidade de uso da forma $\emptyset+V3PP$ encontra-se na faixa 1. Já em Piabas, o uso desta forma é relevante em todas as faixas etárias, tendo peso relativo maior na faixa 3 (0,80).

³⁹ Colocamos o peso relativo referente ao nível 1 da rodada, pois estes grupos, conjuntamente, foram excluídos.

5.1.7 Ø+VINF

Com o menor número de ocorrências, constatamos também o uso da forma $\text{Ø}+\text{VINF}$ pelos falantes das comunidades analisadas. Nas rodadas estatísticas, apenas quatro grupos de fatores foram selecionados como importantes no uso desta forma: forma antecedente, tipo de oração, grau de indeterminação e gênero. Excluímos o grupo tempo e modo verbal, pois já é evidente que com esta variante só vai ocorrer a forma nominal de infinitivo. Os grupos tipo de verbo, preenchimento do sujeito, mudança/manutenção, faixa etária e localidade não foram considerados relevantes pelo VARBRUL, sendo excluídos do arquivo de condições. Nos grupos selecionados, tivemos ainda que fazer algumas alterações, como: 1) amalgamamos as variantes indeterminação parcial com referência implícita no contexto com indeterminação parcial com referência explícita no contexto numa única variante e damos o nome de indeterminação parcial; 2) excluímos as formas pronominais *você(s)*, *a gente*, *ele(s)* e *nós*, tanto com sujeito explícito quanto implícito, e a forma $\text{Ø}+V+SE$, por não apresentarem dados, e juntamos as forma $\text{Ø}+V3PS$ com $\text{Ø}+V3PP$, no grupo forma antecedente.

Os grupos de fatores selecionados foram:

- 1º Forma antecedente
- 2º Tipo de oração
- 3º Grau de indeterminação
- 4º Gênero

5.1.7.1 Forma antecedente

O grupo forma antecedente foi o primeiro a ser selecionado pelo programa e os resultados constam na tabela 41.

Tabela 41: Influência da forma antecedente no uso do $\text{Ø}+\text{VINF}$.

Grupo:			
Forma Antecedente	Apl/T	%	P.R.
$\text{Ø}+V3PS + \text{Ø}+V3PP$	19/297	6	0,38
$\text{Ø}+\text{VINF}$	29/52	56	0,94
TOTAL	48/349	14	-

Nível de significância = 0,009

A partir dos resultados desta tabela, confirmamos mais uma vez que “marcas favorecem marcas”. A probabilidade de emprego da forma $\emptyset+VINF$ é mais elevada quando a forma antecedente é também $\emptyset+VINF$, com peso relativo 0,97.

5.1.7.2 Tipo de oração

A variável tipo de oração só foi selecionada como favorecedora para o uso da variante $\emptyset+VINF$. Os resultados estão expostos na tabela 42:

Tabela 42: Influência do tipo de oração no uso do $\emptyset+VINF$.

Grupo:			
Tipo de Oração	Apl/T	%	P.R.
Simples	25/1429	2	0,40
Principal	3/513	1	0,16
Subordinada	81/622	13	0,89
TOTAL	109/2554	4	-

Nível de significância = 0,000

A oração subordinada é a que exhibe a maior probabilidade do grupo: 0,89, sendo este o único tipo de oração que favorece o uso da forma $\emptyset+VINF$. Ao observar esse uso pelos falantes, vimos que essa forma é utilizada nas orações subordinadas adverbiais finais reduzidas de infinitivo, que sempre indicam o objetivo do fato enunciado na oração principal, como aparece no exemplo (96):

(96) Doc.: É... Tão fazeno o quê aí na praça, sabe dizer? Que desmancham aquilo ali, tudo ali.

Inf.: É... Isso aí... Não, **isso aí é... o aterro pra \emptyset fazer o calçamento**, né? (P, 1, 33).

5.1.7.3 Grau de indeterminação

A partir dos resultados da tabela 43, vemos que a indeterminação com a forma $\emptyset+VINF$ tem maior probabilidade de uso quando a indeterminação é completa, apresentando maior índice percentual, 7%, e peso relativo, 0,65. Assim, ao utilizar esta forma, o falante generaliza o referente, incluindo as três pessoas do discurso. A própria forma nominal de infinitivo já é o verbo no máximo de sua indeterminação e generalidade, contribuindo para o seu caráter indeterminador.

Tabela 43: Influência do grau de indeterminação no uso do $\emptyset+VINF$.

Grupo: Grau de Indeterminação	Apl/T	%	P.R.
Indeterminação completa	80/1132	7	0,65
Indeterminação parcial	29/1422	2	0,38
TOTAL	109/2554	4	-

Nível de significância = 0,000

Nos nossos dados, essa forma é usada como exemplificada em (97):

(97) Doc.: Ah! É sempre bom a gente ouvir uma musiquinha, né?

Inf.: É.

Doc.: Sempre bom.

Inf.: É bom [ouvir uma musiquinha] até **para \emptyset distrair**. (B, 4, 60)

Neste exemplo, vemos que ‘para \emptyset distrair’ não corresponde a nenhum tempo, nenhum modo, nem espécie de ação em particular, o informante usa essa forma de modo genérico, referindo-se a qualquer pessoa que ouve música com a finalidade de se distrair.

5.1.7.4 Gênero

O gênero do informante foi o último grupo selecionado como favorecedor para o uso da variante em análise. De acordo com a tabela 44, os homens são os que mais utilizam a forma $\emptyset+VINF$, com peso relativo 0,56.

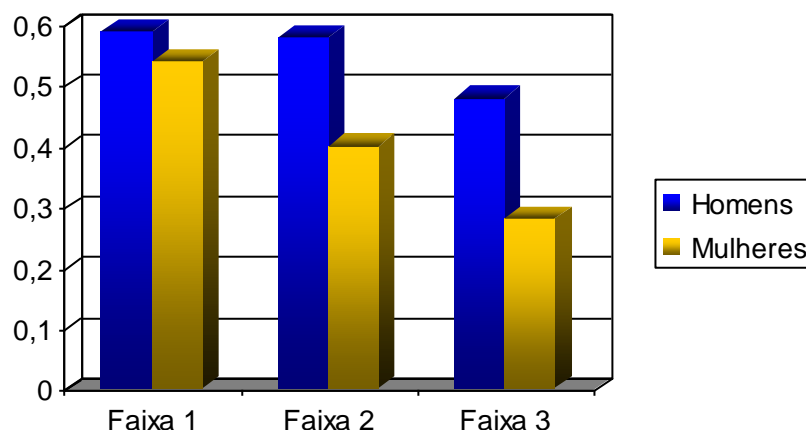
Tabela 44: Influência do gênero no uso do $\emptyset+VINF$.

Grupo: Gênero	Apl/T	%	P.R.
Homens	69/1369	5	0,56
Mulheres	40/1185	3	0,43
TOTAL	109/2554	4	-

Nível de significância = 0,009

Como a faixa etária e a localidade não foram selecionadas pelo programa, resolvemos cruzar esses grupos de fatores com o gênero, pelo VARB2000, para verificar de forma mais específica o comportamento linguístico desses informantes. Quanto ao efeito do cruzamento do gênero com a faixa etária, o gráfico 11 apresenta os resultados.

Gráfico 11: Efeito do gênero e da faixa etária na escolha da forma $\emptyset+VINF$.
P.R.



Nível de significância = 0,033

A partir desse gráfico, vemos que são os homens de todas as faixas etárias que mais utilizam a forma $\emptyset+VINF$ para indeterminar o sujeito, sendo que esse uso vai aumentando na medida que a faixa etária diminui. É possível observar que, entre eles, o uso desta forma é mais significativo nas faixas 1 e 2. Com relação às mulheres, ocorre o mesmo, porém só há valor significativo entre as mulheres da faixa 1, com peso relativo um pouco acima de 0,50. Esta parece ser uma forma inovadora que, provavelmente, está tomando o lugar do $\emptyset+V+SE$.

A fim de observar se o comportamento lingüístico de homens e mulheres se diferencia nas comunidades, fizemos um cruzamento entre o gênero do informante e as localidades e obtivemos os seguintes resultados:

Tabela 45: Efeito da faixa etária e da localidade no uso do $\emptyset+VINF$.

Gênero/ Localidade	Bananal/Barra dos Negros			Piabas		
	Nº/T	%	P.R.	Nº/T	%	P.R.
Homens	42/913	5	0,53	27/456	6	0,60
Mulheres	22/857	3	0,38	18/328	5	0,57
TOTAL	64/1770	4	-	45/784	6	-

Observamos que em Piabas a forma $\emptyset+VINF$ é usada de modo significativo tanto entre os homens quanto entre as mulheres, com peso relativo 0,60 e 0,57, respectivamente. Em Bananal / Barra dos Negros seu uso é significativo apenas entre os homens (0,53). Se observarmos os percentuais, vemos que esta forma é mais usada em Piabas.

5.2 ANÁLISE COMPARATIVA

Dentre as variantes examinadas nesta pesquisa, os grupos de fatores apresentados no quadro 2 foram considerados relevantes para aplicação de cada uma das formas de indeterminação, de acordo com a ordem de seleção.

Quadro 1: Posição dos grupos de fatores selecionados para cada variante da indeterminação.

GRUPO DE FATORES	A gente	Nós	Você	Eles	Ø+V3PS	Ø+V3PP	Ø+VINF
Forma Antecedente	1°	1°	1°	1°	1°	2°	1°
Grau de Indeterminação	2°	6°	2°	2°	3°	3°	3°
Tempo e Modo Verbal	3°	3°	4°	3°	2°	1°	-
Tipo de Verbo	-	7°	-	-	-	-	-
Preenchimento do Sujeito	-	2°	-	-	-	-	-
Tipo de Oração	-	-	-	-	-	-	2°
Mud./Man. do Referente	-	-	-	-	-	-	-
Gênero	6°	-	5°	-	4°	-	4°
Faixa Etária	4°	-	3°	-	5°	-	-
Faixa Etária + Gênero	-	4°	-	4°	-	-	-
Localidade	5°	5°-	-	-	-	4°	-

Como podemos ver, alguns grupos de fatores selecionados foram os mesmos para todas as variantes da indeterminação, como por exemplo, a forma antecedente, o grau de indeterminação e o tempo e modo verbal. As variáveis sociais gênero e faixa etária também foram selecionadas quer separadas, com as formas *a gente*, *você*, $\text{Ø}+V3PS$ e $\text{Ø}+VINF$, quer combinadas, com as formas *nós* e *eles*. A variável localidade só foi significativa no uso de *a gente*, *nós* e $\text{Ø}+V3PP$. Outros grupos foram selecionados em apenas uma das variantes e o grupo mudança/manutenção do referente não foi selecionado como favorecedor a aplicação da regra para nenhuma das variantes da indeterminação analisadas neste trabalho.

A forma antecedente mostrou-se, em nossa investigação, o grupo mais significativo, sendo o primeiro a ser selecionado como favorecedor para todas as variantes, com exceção apenas da variante $\text{Ø}+V3PP$, para a qual foi selecionado em segundo lugar. Os resultados numéricos confirmaram a hipótese de que a ocorrência de uma determinada forma tende a desencadear uma série de repetições, em estruturas paralelas. Ao compararmos este resultado com os resultados de Cunha (1993), é possível observar que este grupo ocupa também a primeira ou a segunda posição para todas as variantes por ela analisadas.

Ao cruzar o grupo forma antecedente com a mudança/ manutenção do referente, os números mostraram que a mesma forma é sempre usada quando a referência é também a mesma. A partir do momento que se muda a referência, a forma também é mudada. O mesmo resultado podemos ver também em Cunha.

O grau de indeterminação também foi considerado importante no uso das variantes, sendo selecionado por todos os grupos ora nas primeiras, ora nas últimas posições. A partir deste grupo podemos perceber o valor de cada variante. *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 mais genérica, podendo referir-se a qualquer pessoa do discurso.

O tempo e modo verbal foi também selecionado por todas as variantes e dentre as quatro primeiras posições mais significativas, com exceção da forma $\emptyset+VINF$, com a qual o uso da forma nominal de infinitivo é categórico, já que o uso desta variante se dá apenas com o infinitivo. De modo geral, vimos que o tempo e modo verbal mais utilizado pelos falantes é o presente do indicativo. Isso é detectado também na análise de Godoy (1999) e de Setti (1997). Nesta pesquisa, o presente do indicativo é o mais utilizado, em número de ocorrências, com todas as variantes da indeterminação, mas em probabilidade, este foi selecionado para o uso de *a gente*, *você*, *eles* e $\emptyset+V3PS$. Em segundo lugar, são usados os tempos do pretérito, também do indicativo. Seu uso teve relevância com as formas *nós* e $\emptyset+V3PP$.

O tipo de verbo e o preenchimento do sujeito só foram escolhidos para o uso da variante *nós*, parecendo não ser muito significativo para o uso das variantes da indeterminação. O tipo de oração só foi selecionado para o uso do $\emptyset+VINF$, pois o uso dessa variante está condicionado às orações subordinadas adverbiais finais reduzidas de infinitivo, como ficou comprovado nos dados estatísticos.

Os fatores extralingüísticos atuaram significativamente na escolha das formas de indeterminação ora isolados, ora conjugados com outro fator.

Na amostra total, os resultados referentes ao fator gênero mostraram que homens e mulheres têm comportamentos diferenciados na escolha dos recursos para indeterminar o sujeito. Estas só se mostraram relevantes no uso da forma pronominal *a gente*, enquanto que os homens utilizam mais as formas *você*, $\emptyset+V3PS$, $\emptyset+VINF$. Quando fizemos as rodadas com as formas *nós*, *eles* e $\emptyset+V3PP$, como valor de aplicação, o fator gênero não foi selecionado como fator significativo, pois estas variantes são usadas com relativa igualdade entre homens e mulheres.

Em relação à faixa etária, há também um comportamento diferenciado entre falantes das três faixas. A partir dos resultados, vimos que, na amostra total, a faixa 1 não favorece o uso dos recursos de indeterminação, a sua probabilidade fica sempre próxima de 0,50 para qualquer variante. São nas faixas 2 e 3 que estão as maiores probabilidades de uso. *A gente* e *você* são mais usadas pelos informantes da faixa 2 e $\emptyset+V3PS$ pelos da faixa 3. A faixa etária não foi considerada significativa para o uso das variantes *nós*, *eles*, $\emptyset+V3PP$ e $\emptyset+VINF$.

Em algumas rodadas, o grupo gênero só foi selecionado pelo programa combinado com a faixa etária, nesses casos, foram selecionados para o uso das formas *nós* e *eles*. Assim, as mulheres da faixa 3 utilizam muito mais a forma *nós* e os homens da faixa 3 e as mulheres da faixa 1 utilizam a forma *eles*.

Em relação às localidades, ambas apresentam comportamentos lingüísticos diferenciados quanto ao uso dos recursos de indeterminação. Piabas mostrou-se mais conservadora, com um uso mais intenso das formas *nós* e $\emptyset+V3PP$. Enquanto que os falantes de Bananal / Barra dos Negros utilizam mais a forma *a gente*, mostrando-se mais inovadora. Piabas, apesar de fazer parte de um município de fundação mais recente do que o município a que pertence às comunidades de Bananal / Barra dos Negros, mantém características conservadoras, visto que está situada dentro da caatinga, numa localização onde não é passagem nem entroncamento rodoviário de viajantes, além de não ter atrativo algum. Lá, os moradores saem para outras localidades, mas ninguém de outras localidades vai lá. As mulheres vão para as feiras e para atendimento médico no próprio município de Caém e nas cidades vizinhas, como Jacobina, e Campo Formoso, e algumas vezes vão ao estado de São Paulo visitar parentes. E os homens vão a São Paulo a procura de emprego. Bananal / Barra dos Negros fazem parte de um município que atrai muita gente devido ao turismo da região, havendo, assim, mais contatos entre diferentes grupos populacionais, o que pode estar influenciando na fala dessas comunidades.

6 CONSIDERAÇÕES FINAIS

Conforme mencionamos na introdução deste trabalho, propusemo-nos mostrar que existem outras formas para se indeterminar o sujeito em português que não são abordadas pela GT; além disso, propusemo-nos também verificar quais são os fatores lingüísticos e extralingüísticos que condicionam o uso das formas de indeterminação utilizadas pelos falantes de comunidades rurais do semi-árido baiano.

De fato, os resultados mostram que há uma grande distância entre o que é colocado nas Gramáticas Tradicionais e o que ocorre na língua falada de comunidades rurais, no que se refere à indeterminação do sujeito. Primeiramente, constatamos que está em uso um número muito maior de formas de indeterminação do sujeito do que descreve a tradição gramatical, que contempla apenas o uso do verbo na terceira pessoa do plural e o uso do verbo na terceira pessoa do singular mais o pronome “se”. Os resultados também mostraram que as formas apresentadas pela Gramática Tradicional não são as mais utilizadas nos *corpora*, as formas *a gente* e $\emptyset + V3PS$ são as mais empregadas.

Pudemos também confirmar que a indeterminação é um fenômeno lingüístico explicável essencialmente nos níveis semântico e discursivo, visto que o que se indetermina é a referência do sujeito, estando esta essencialmente relacionada ao contexto. Assim, mesmo quando o sujeito é expresso por um pronome, pode ser considerado como um recurso de indeterminação, se não houver possibilidade de recuperação de sua referência no contexto.

A indeterminação também comporta graus, pois os recursos usados pelo falante para indeterminar o sujeito não agem da mesma forma, o conteúdo genérico de cada recurso tem características referenciais diferenciadas. Na sua realização eles apresentam uma capacidade de generalização que permite uma grande oscilação de seus usos indeterminados. Assim, a indeterminação cresce à medida que deixam de existir elementos no contexto que permitam a interpretação de qual seja seu referente. Desta forma, o grau de indeterminação será maior ou menor de acordo com a dependência da interpretação da forma indeterminadora em relação ao contexto.

Finalmente, a indeterminação mostrou-se um fenômeno sensível a condicionamentos lingüísticos e extralingüísticos. Em relação aos primeiros, a forma antecedente, o grau de indeterminação e o tempo verbal são os mais importantes, sendo selecionados como favorecedores o uso de todas as variantes da indeterminação. Em relação aos extralingüísticos, o gênero e a faixa etária são os mais importantes, sendo selecionados ora de forma isolada, ora conjuntamente.

Em síntese, destacamos as tendências quanto ao uso dos recursos de indeterminação do sujeito no português rural do semi-árido baiano.

- 1) A forma antecedente mostrou-se o grupo mais significativo, sendo selecionado como favorecedor o uso de todas as variantes da indeterminação. Sendo assim, quando o falante opta por usar um recurso de indeterminação, numa seqüência discursiva, a tendência é o desencadeamento de uma série de repetições da forma empregada na primeira referência. Entretanto, quando o falante muda o referente, a forma escolhida também se altera.
- 2) Quanto ao tempo verbal, os tempos do indicativo são os mais freqüentes, principalmente o presente. Este fato está diretamente ligado ao contexto discursivo dos falantes.
- 3) Em relação ao grau, há uma diferenciação no uso das formas de indeterminação, o falante utiliza *você*, $\emptyset+V3PS$, $\emptyset+V3PP$ e $\emptyset+VINF$ quando a indeterminação é completa, não havendo possibilidade de determinação do referente no contexto discursivo. Utiliza as formas *nós* ou a *gente* quando a indeterminação é parcial com referência implícita, pois usando estas formas, de modo geral, o falante está se referindo a ele mesmo mais as pessoas da comunidade. A forma *eles* é utilizada quando a indeterminação é parcial com referência explícita, pois ao usar esta forma, é possível identificar, através de elementos do texto, o grupo a que o falante está se referindo.
- 4) Quanto ao gênero, este só é significativo para o uso de *você*, *a gente*, $\emptyset+V3PS$ e $\emptyset+VINF$ sendo que as mulheres utilizam mais a forma *a gente* do que os homens. Estes tem um uso mais freqüente das formas *você*, $\emptyset+V3PS$ e $\emptyset+VINF$.
- 5) Quanto à faixa etária, na amostra geral, a faixa 1 não utiliza de forma significativa as formas de indeterminação, são nas faixas 2 e 3 que estão os usos mais intensos, sendo que na faixa 2 os falantes utilizam mais as formas *a gente* e *você* e na faixa 3, a forma $\emptyset+V3PS$.

- 6) No que diz respeito à localidade, esta só foi selecionada como significativa pelo programa para o uso de *a gente*, *nós* e $\emptyset+V3PP$. Em Bananal / Barra dos Negros, os falantes utilizam mais a variante *a gente* (considerada inovadora enquanto pronome e também como forma para indeterminar o sujeito) e em Piabas as formas *nós* e $\emptyset+V3PP$ (consideradas conservadoras, a primeira enquanto pronome e a segunda enquanto forma para indeterminar o sujeito). Parece-nos que o uso destas variantes reflete características de ambas as comunidades.
- 7) Essas comunidades apresentam também comportamentos lingüísticos diferenciados entre os gêneros e as faixas etárias quando vistas de forma isolada. Em vista disso, ao fazer rodadas isolando cada comunidade e cruzando com os grupos gênero e faixa etária, obtivemos os seguintes resultados:
- a. Quanto à localidade combinada com o gênero, a forma *nós* é utilizada de forma significativa entre homens e mulheres de Piabas, com favorecimento maior para as mulheres (cf. tabela 16); a forma *a gente* tem um uso significativo apenas entre as mulheres de Bananal / Barra dos Negros, em Piabas não é significante o uso desta variante. *Você* é mais usado pelos homens de Bananal / Barra dos Negros e pelas mulheres de Piabas, porém estes resultados não refletem exatamente a realidade lingüística dessas localidades, visto que: a) houve uma alta incidência de uso desta forma por um único informante de Bananal / Barra dos Negros; e b) o número de ocorrências desta forma em Piabas é muito pequeno: são apenas 13 casos, sendo 9 de usos das mulheres e 4 dos homens (cf. tabela 25). A forma $\emptyset+V3PS$ é usada tanto pelos homens de Piabas como pelos de Bananal / Barra dos Negros (cf. tabela 33). As formas *eles*, $\emptyset+V3PP$ e $\emptyset+VINF$ não foram selecionadas pelo programa nestas rodadas.
 - b. Quanto à localidade combinada com a faixa etária, a forma *a gente* é usada pelos falantes das faixas 1 e 2 (sendo esta última de forma mais significativa) de Bananal / Barra dos Negros; a forma *nós* é usada pelos falantes de todas as faixas etárias de Piabas, principalmente da faixa 3 (cf. tabelas 17). *Você* é usado mais intensamente na faixa 2 de Bananal / Barra dos Negros (isso devido a um único informante, como já foi explicado) e na faixa 3 de Piabas. A forma $\emptyset+V3PS$ é

mais usada na faixa 3 de Bananal / Barra dos Negros e tem seu uso próximo do ponto neutro em todas as faixas etárias de Piabas.

- 8) Não foi possível realizar uma análise estatística com a variante $\emptyset+V+SE$, devido à pequena quantidade de ocorrências, apenas 15, o que impossibilitou a análise da regra variável.

Os resultados obtidos ratificam a necessidade de se diminuir a distância existente entre a prática pedagógica e a utilização efetiva da língua. Estes podem fornecer elementos para a compreensão dos usos das formas utilizadas pelos falantes para indeterminar o sujeito, possibilitado, desse modo, o estabelecimento de práticas pedagógicas mais condizentes com a realidade circundante.

Esta abordagem da indeterminação não contempla todas as nuances deste fenômeno numa comunidade rural, até porque, diante de uma questão tão complexa como a indeterminação, nossa intenção não foi a de propor uma abordagem exaustiva e definitiva. Mas cremos que os resultados apresentados podem servir de ponto de partida para a elaboração de trabalhos que se aprofundem em outros níveis. Além disso, sugerem a importância de continuidade e aprofundamento da análise de alguns fatores lingüísticos e extralingüísticos em pesquisas futuras, como, por exemplo, refletir características discursivas.

REFERÊNCIAS

- AGUILERA, Vanderci de Andrade. *Atlas lingüístico do Paraná*. Curitiba: Imprensa Oficial do Estado, 1994.
- ALBÁN, Maria Del Rosário; FREITAS, Judith. Eu, você et alia em três diálogos. *Estudos Lingüísticos e Literários*. Salvador, n. 11, p. 25-38, 1991a.
- ALBÁN, Maria Del Rosário; FREITAS, Judith. Nós ou a gente? *Estudos Lingüísticos e Literários*. Salvador, n. 11, p. 75-90, 1991b.
- ALKMIM, Tânia Maria. Considerações sobre o campo da Sociolingüística. In: ALBANO, Eleonora et al. (org.). *Saudades da língua: a lingüística e os 25 anos do Instituto de Estudos da Linguagem da UNICAMP*. Campinas, SP: Mercado de Letras, 2003b.
- ALKMIM, Tânia Maria. Sociolingüística: parte I. In: MUSSALIN, Fernanda; BENTES, Anna Cristina (org.). *Introdução à lingüística: domínios e fronteiras*. São Paulo: Cortez, 2003a. Vol 1.
- ALMEIDA, Maria Lúcia Leitão de. *Sujeito indeterminado na fala*. 1992. 283f. Tese (Doutorado em Lingüística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1992.
- ALMEIDA, Napoleão Mendes de. *Gramática metódica da língua portuguesa*. 44 ed. São Paulo: Saraiva, 1999.
- ALMEIDA, Norma Lúcia F.e CARNEIRO, Zenaide Oliveira. N. (Org.). *Amostras da língua falada em Bananal/Barra dos Negros* (Rio de Contas). Feira de Santana: UEFS, Departamento de Letras e Artes, 2002b. (Coleção Amostras da fala do Semi-Árido - Vol. II).
- ALMEIDA, Norma Lúcia F.e CARNEIRO, Zenaide Oliveira. N. (Org.). *Amostras da língua falada em Piabas* (Anselino da Fonseca). Feira de Santana: UEFS, Departamento de Letras e Artes, 2002a. (Coleção Amostras da fala do Semi-Árido - Vol. I).
- ARAGÃO, Maria do Socorro; MENEZES, Cleuza Bezerra de. *Atlas Lingüístico da Paraíba*. I-II. Brasília: UFPB/CNPq, Coordenação Editorial, 1984.
- BECHARA, Evanildo. *Moderna gramática portuguesa*. 37 ed. Rio de Janeiro: Lucerna, 1999.
- BECHARA, Evanildo. *Moderna gramática portuguesa: cursos de 1º e 2º graus*. 34 ed. São Paulo: Companhia Editora Nacional, 1992.
- BENVENISTE, Émile. *Problemas de lingüística geral I*. Tradução Maria da Glória Novak e Maria Luisa Néri; rev. Prof. Isaac Nicolau Salum. 4 ed. Campinas, SP: Pontes, 1995.
- BRESCANCINI, Cláudia Regina. A análise de regra variável e o programa VARBRUL 2S. In: BISOL, Leda e BRESCANCINI, Cláudia (Org.). *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, 2002.

BRIGHT, William. Sociolinguistics. In: BRIGHT, William (Org.). *Proceeding of the UCLA sociolinguistics conference, 1964*. 3 ed. Monton: The Hauge, 1966. p. 11-15.

CARDOSO, Suzana Alice M. A dialetologia no Brasil: perspectivas. São Paulo, *DELTA*, n. especial, v. 15, p. 233-255, 1999.

CARDOSO, Suzana Alice M. Dialetologia: trilhas seguidas, caminhos a perseguir. São Paulo, *DELTA*, n. especial, v. 17, p. 25-44, 2001.

CARDOSO, Suzana Alice M. Perspectivas para a dialectologia no Brasil. In: CARDOSO, Suzana Alice M. (Org.). *Diversidade lingüística e ensino*. Salvador: EDUFBA, 1996. p. 105-112.

CARDOSO, Suzana Alice M.; MOTA, Jacyra Andrade. Um passo da geolingüística brasileira: o Projeto AliB. IN: RONCARATI, Cláudia; ABRAÇADO, Jussara. *Português brasileiro: contato lingüístico, heterogeneidade e história*. Rio de Janeiro: 7Letras, 2003. p. 39-49.

CAVALCANTE, Silvia Regina de Oliveira. *A indeterminação do sujeito na escrita padrão: a imprensa carioca nos séculos XIX e XX*. 1999. 115 f. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999.

CAVALCANTE, Silvia Regina de Oliveira. Formas de indeterminação na imprensa carioca dos séculos XIX e XX. In: ALKMIM, Tânia (Org.). *Para a história do português brasileiro*. São Paulo: Humanitas/ FFLCH/ USP, 2002. vol. 3.

CAVALCANTE, Vilma Maria Reis. *A indeterminação do sujeito no português oral culto de Fortaleza – CE*. 1998. 108f. Dissertação (Mestrado em Lingüística). Universidade Federal do Ceará, Fortaleza, 1998.

CEDERGREN, Henrietta; SANKOFF, David. Variable rules: performance as a statistical reflection of competence. *Language*. n. 50, p. 333-355, 1974.

CEGALLA, Domingos Paschoal. *Novíssima gramática da língua portuguesa*. 39 ed. São Paulo: Nacional, 1995.

CENSO 2000. Disponível em: <http://www.ibge.gov.br/home/estatistica/censo2000>. Acesso em: 05 dez. 2005.

CHAMBERS, J. K.; TRUDGILL, Peter. *La dialectología*. Tradução Carmen Morán González. Madri: visor Libros, 1994.

CINTRA, Luis Felipe L. *Sobre “Formas de Tratamento” na língua portuguesa*. Lisboa: Livros Horizonte, 1972.

COUTINHO, Ismael de Lima. *Pontos de gramática histórica*. 4 ed. Rio de Janeiro: Livraria Acadêmica, 1958.

CUNHA, Celso; CINTRA, Luís F. Lindley. *Nova gramática do português contemporâneo*. 3 ed. Rio de Janeiro: Nova Fronteira, 2001.

CUNHA, Cláudia de Souza. *Indeterminação pronominal do sujeito*. 1993. 111f. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1993.

DUARTE, Maria Eugenia L. *A perda do princípio 'Evite Pronome' no português brasileiro*. 1995. 151f. Tese (doutorado em Lingüística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, SP, 1995.

FARACO, Carlos Alberto. *Lingüística histórica: uma introdução ao estudo das línguas*. São Paulo: Parábola Editorial, 2005.

FARACO, Carlos Alberto. *O tratamento você em português: uma abordagem histórica*. Fragmenta. Curitiba: UFPR, n. 13, 1996.

FERNANDES, Eulália; GORSKY Edair. A concordância verbal com os sujeitos Nós/ A gente: um mecanismo do discurso em mudança. In: *Atas do I Simpósio sobre a Diversidade Lingüística no Brasil*. Salvador: Instituto de Letras da UFBA. p. 183-7, 1986.

FERREIRA Carlota; CARDOSO, Suzana Alice M. *A dialetologia no Brasil*. São Paulo: Contexto, 1994.

FERREIRA, Carlota *et al.* A pessoa e a não-pessoa em discursos de informantes do Projeto NURC/Salvador. *Estudos Lingüísticos e Literários*. Salvador, n. 11, p. 39-52, 1991.

FERREIRA, Carlota *et al.* *Atlas Lingüístico de Sergipe*. Salvador: Universidade Federal da Bahia / Fundação Estadual de Cultura de Sergipe, 1987.

FERREIRA, Carlota; CARDOSO, Suzana Alice. *A dialetologia no Brasil*. São Paulo: Contexto, 1994. (Repensando a língua portuguesa).

FERREIRA, Mauro. *Aprender e praticar gramática*. São Paulo: FTD, 1992.

FREITAS, Judith. Nós e a gente em elocuições formais. *Estudos Lingüísticos e Literários*. Salvador, n. 11, p. 91-102, 1991.

GILLIÉRON, Jules; EDMONT, Edmond. *Atlas Linguistique de la France (ALF)*. Paris: Honoré Champion, 1902-1910.

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

IKEDA, Sumiko Nishitani. A função do se. *Cadernos da PUC*. São Paulo, n. 5, p. 111-147, 1980.

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

- INFANTE, Ulisses. *Curso de gramática aplicada aos textos*. São Paulo: Scipione, 1995.
- JABERG, Karl; JUD, Jakob. *Spraach - und Sachatlas Italiens und der Südschweiz, I-VIII*. Zofingen: Rieger & Co, 1928-1940.
- KURATH, Hans *et al.* *Linguistic Atlas of New England (LANE)*. Brown University Press, 1939-1943. 3 vol.
- KURY, Adriano da Gama. *Gramática fundamental da língua portuguesa*. São Paulo: LISA, 1989.
- LABOV, William. *Modelos sociolingüísticos*. Traducción y versión española de Júlio Borrego Nieto. Madrid: Ediciones Cátedra, 1983.
- LABOV, William. *The social stratification of english in New York city*. Washington, D. C.: Center for Applied Linguistics, 1966.
- LAVANDERA, Beatriz R. *Variación y significado*. Buenos Aires: Hachette, 1984.
- LOPES, Célia Regina dos Santos. *Nós e A gente no português falado culto do Brasil*. 1993. 140 f. Dissertação (Mestrado em Lingüística)- Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1993.
- LUFT, Celso Pedro. *Moderna gramática brasileira*. 13 ed. São Paulo: Globo, 1996.
- MACHADO, Márcia dos Santos. *Sujeitos pronominais 'nós' e 'a gente': variação em dialetos populares do norte fluminense*. 1995. 252 f. Dissertação (Mestrado em Lingüística)- Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1995.
- MELHORAMENTOS. *Dicionário prático da língua portuguesa*. 10 ed. São Paulo: Melhoramentos, 1995.
- MELO, Gladstone Chaves de. *Gramática fundamental da língua portuguesa*. 4 ed. Rio de Janeiro: Ao Livro Técnico, 1980.
- MELO, Gladstone Chaves de. *Novo manual de análise sintática*. 3 ed. Rio de Janeiro: Livraria Acadêmica, 1967.
- MENON, Odete Pereira da Silva. *Analyse sociolinguistique de l'indétermination du sujets dans lê portugais parlé au Brésil, à partir des données du NURC/SP*. 1994. 397f. Tese (Doutorado em Lingüística). Université de Paris VII, Paris, 1994.
- MENON, Odete Pereira da Silva. O sistema pronominal do português do Brasil. *Letras*, Curitiba, n. 44, p. 91-106, 1995.
- MICHELETTI, Helena e FRANCHETTI, Susan Klein. A indeterminação do sujeito: um estudo sociolingüístico. *Estudos Lingüísticos: anais de seminários do GEL*. Taubaté, SP, v. XXV, p.629-635, 1996.

MILANEZ, Wânia. *Recursos de indeterminação do sujeito*. 1982. 143 f. Dissertação (Mestrado em Lingüística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1982.

MOLLICA, Maria Cecília. Fundamentação teórica: conceituação e delimitação. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (Org.). *Introdução à sociolingüística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 9-14.

MOLLICA, Maria Cecília; BRAGA, Maria Luiza (Org.). *Introdução à sociolingüística: o tratamento da variação*. São Paulo: Contexto, 2003.

MONTEIRO, José Lemos. *Para compreender Labov*. 2 ed. Petrópolis, RJ: Vozes, 2000.

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

MOTA, Jacira Andrade. O atlas lingüístico do Brasil (AliB) e os estudos geolingüísticos no Brasil. Londrina, *SIGNUM: estudos da linguagem*, n. 5, p. 365-374, dez. 2002b.

MOTA, Jacyra Andrade. *Atlas lingüístico do Brasil: aspectos metodológicos*, 2002a. mimeo.

NARO, Anthony J. Modelos quantitativos e tratamento estatístico. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). *Introdução à sociolingüística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 15-25.

NARO, Anthony J. The genesis of the reflexive impersonal in Brazilian portuguese: a study in syntatic change as a surface phenomom. *Language*, n. 52, p. 779-810, 1976.

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

OLIVEIRA, Dercir Pedro de; ISQUERDO, Aparecida Negri. A nova dialetologia: investigações e resultados. In: RONCARATI, Cláudia; ABRAÇADO, Jussara. *Português brasileiro: contato lingüístico, heterogeneidade e história*. Rio de Janeiro: 7Letras, 2003. p. 50-54.

OMENA, Nelize Pires de A referência variável da primeira pessoa do discurso no plural. In: NARO, Anthony, *et alii*. *Relatório Final de Pesquisa: Projeto Subsídios do Projeto Censo à Educação*. Rio de Janeiro, UFRJ, v. 2, p. 286-319, 1986.

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

PAIVA, Maria da Conceição; DUARTE, Maria Eugênia L. (Org.). *Mudança lingüística em tempo real*. Rio de Janeiro: Contra Capa, 2003.

PERINI, Mário Alberto. *Gramática descritiva do português*. 4 ed. São Paulo: contexto, 2004.

PINTZUK, Susan. *VARBRUL programs*. Trad.de Ivone Isidoro Pinto, revisão de Maria Thereza Gomes Fioreti e Maria Marta Pereira Scherre (Coord.). Original inglês. Inédito, 1988.

RADTKE, Edgard; THUN, Harald. Nuevos caminos dela geolingüística românica: un balance. In: RADTKE, Edgard; THUN, Harald. *Neue Wege der Romanischen Geolinguistik*. Akten des Symposiums zur Empirischen Dialektologie. Kiel: Westensee, 1996. p. 25-49.

RAGO, Margareth. Descobrimdo historicamente o gênero. *Cadernos Pagu*, n. 11, p. 89-98, 1998.

RIBEIRO, José *et al.* *Esboço de um atlas lingüístico de Minas Gerais*, I. Rio de Janeiro: Fundação Casa de Rui Barbosa/ Universidade Federal de Juiz de Fora, 1977.

ROCHA LIMA, Carlos Henrique da. *Gramática normativa da língua portuguesa*. 37 ed. Rio de Janeiro: José Olympio, 1998.

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

ROSSI, Nelson. Dialectologia. In: Houaiss. Antonio (ed.). *Enciclopédia Mirador Internacional*, São Paulo: Melhoramentos, v.7, p. 3298-3304, 1980.

ROSSI, Nelson; FERREIRA, Carlota; ISENSEE, Dinah Maria. *Atlas prévio dos falares baianos*. Rio de Janeiro: INL-MEC, 1963.

ROUSSEAU Pascale, & SANKOFF, David. Advances in variable rule methodology. In: ROUSSEAU Pascale, & SANKOFF, David (Org.). *Linguistic variation: models and methods*. New York: Academic Press, 1978. p. 57-69.

SAID ALI, Manuel. *Gramática histórica da língua portuguesa*. 6 ed. São Paulo: Melhoramentos, 1966.

SANKOFF, David. Variable rules. In: AMMON, UUlrich; DITTMAR, Norbert; MATTHEIER, Klaus J. (ed.). *Sociolinguistics: aAn international handbook of the science of language and society*. New York: Walter de Gruyter, 1988. p. 984-998.

SCHERRE, Maria Marta Pereira. A concordância de número nos predicativos e participios passivos. *Organon*. Porto Alegre, v. 5, n. 18, 1991.

SCHERRE, Maria Marta Pereira. *Introdução ao pacote VARBRUL para microcomputadores*. UFRJ/UnB, 1992/1993. Inédito.

SCHERRE, Maria Marta Pereira. Paralelismo formal e cognição. *Boletim da Associação Brasileira de Lingüística*. São Paulo, USP, n. 13, p. 43-53, dez. 1992.

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

SETTI, Adriane Cristina Ribas. *A indeterminação do sujeito nas três capitais do sul*. 1997. 116 f. Dissertação (Mestrado em Lingüística). Universidade Federal do Paraná, Paraná, 1997.

SILVA, Giselle Machine de Oliveira; PAIVA, Maria da Conceição. Visão de conjunto das variáveis sociais. In: SILVA, Giselle Machine de Oliveira; SCHERRE, Maria Marta Pereira. *Padrões sociolingüísticos*. 2 ed. Rio de Janeiro: Tempo Brasileiro, 1998.

TARALLO, Fernando. *A pesquisa sociolingüística*. 3 ed. São Paulo: 1990. (Série Princípios).

TEIXEIRA, Eliana Sandra Pitombo. *Era uma vez você*. 2001. 167 f. Tese (Doutorado em Lingüística). Instituto de Letras. Universidade Federal da Bahia, Salvador, 2001.

TRAVAGLIA, Luiz Carlos. *Gramática e interação: uma proposta para o ensino de gramática no 1º e 2º graus*. 8 ed. São Paulo: Cortez, 2002.

TUPINÁ, Heloísa Marques. Abrangência pessoal dos processos de indeterminação do agente. *Alfa*, São Paulo, n. 28, p. 63-69, 1984.

WEINREICH, Uriel; LABOV, William. e HERZOG, Marvin. Empirical foundations for a theory of language change. In: LEHMANN, W. e MALKIEL, Y. (ed.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p. 95-188.