

Universidade Federal de Minas Gerais

**Variação em itens lexicais
terminados em /l/+vogal
na região de Itaúna/MG**

Alan Jardel de Oliveira
2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Alan Jardel de Oliveira

Variação em itens lexicais terminados em /l/+vogal na região de Itaúna/MG

Dissertação apresentada ao Curso de Pós-Graduação em Estudos Lingüísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Lingüística.

Área de Concentração: Lingüística
Linha de Pesquisa: B – Estudo da Variação e Mudança Lingüística
Orientadora: Prof^a Dr^a Maria do Carmo Viegas

Belo Horizonte
Faculdade de Letras da UFMG
2006

**Dissertação aprovada em 30/03/2006 pela Banca Examinadora
constituída pelos Professores Doutores**

Profª Drª Maria do Carmo Viegas – UFMG
Orientadora

Prof. Dr. Marco Antônio de Oliveira – PUC/MG

Prof. Dr. Iuri da Costa Leite – FIOCRUZ/RJ

Dedicatória

Aos meus pais.

Agradecimentos Especiais

Ao Marcelo,

Por tudo que ele me fez descobrir e que descobrimos juntos nesses últimos tempos e, especialmente, pela sua paciência. Agradeço pelos dias incontáveis que passamos juntos, conversando sobre estatística. Agradeço pela dedicação, pela preocupação, pela força e por tudo mais que ele fez por mim e por este trabalho. Sem ele certamente grande parte deste trabalho não existiria.

À Maria do Carmo,

Pela motivação, pela dedicação, pela paciência, pela confiança depositada em mim, por ter me orientado de forma fantástica, deixando-me ir aonde eu quisesse, mesmo sabendo que, em algumas vezes, o caminho não levava para muito longe ou levava para longe demais. Agradeço pelo incentivo nos momentos mais complicados e por ter me apoiado quando decidi entrar também no ‘mundo dos números’. Agradeço por tudo que pude aprender com ela: a paciência na análise, a rigidez metodológica, a importância que se deve dar àquilo que de fato foi observado e a visão desprendida de preconceitos frente àquilo que é diferente do que achamos ser o melhor caminho.

Agradecimentos

À Dani,

Pela sua presença, pela motivação, pelas suas sugestões e por participar desse processo ‘doloroso’ junto comigo;

À minha família – Cássia, minha mãe; Dalmo, meu pai; meus irmãos Alain, Almir, Alex e Marinalva e meus sobrinhos Gabriel e Rafael – pelo apoio e por estarem sempre presentes;

Ao Marcelo e à Laninha, por me acolherem sempre de um jeito que só eles sabem fazer;

À Tia Bê e à Tia Neide, por terem me ajudado tanto nos períodos em que estive em Itaúna;

Aos colegas, sem os quais esse período não teria tanta graça: Kátia, Daniela, Rubens, Léo, Raquel, Daniel, Sandro e Hilda. Agradecimentos especiais ao Sandro, pelas leituras e sugestões, e à Hilda;

Ao Alex Manzi, pela disposição em quase ler este trabalho e pela compreensão neste período de tempo escasso;

A todos os meus amigos e familiares, por estarem sempre comigo; entendendo, ou não, minhas ausências;

Ao Iuri, pelas contribuições que deu a este trabalho e pela disponibilidade em ajudar;

Ao Prof. Marco Antônio de Oliveira, por ter sido o responsável pelo meu interesse no estudo da variação lingüística. Agradeço também pelas excelentes conversas e pelas contribuições que ele deu a este trabalho;

Ao Prof. Gregory Guy, por ter sido o responsável pelo meu interesse no aprofundamento dos métodos de análise quantitativa;

À Prof^a Thaís Cristófaró, pelas sugestões, pelas indicações bibliográficas e pela disponibilidade;

À Prof^a Cibele Comini, pela disponibilidade, pelas leituras e pelas sugestões que deu a este trabalho;

Aos professores que tão atenciosamente responderam minhas dúvidas enviadas por e-mail: Maurílio Vieira, Tony Sardinha, José Olímpio, Susan Pintzuk, Sali Tagliamonte, John Paolillo e Gregory Guy;

Aos professores das disciplinas cursadas: Eunice Nicolau, Jânia Ramos, José Olímpio, César Reis, Maria do Carmo, Thaís Cristófaró e Eleonora Albano;

Aos informantes, sem os quais este trabalho não teria sido possível;

Aos jornais de Itaúna, por disponibilizarem o material para o *corpus* escrito;

Aos funcionários do Poslin e da Faculdade de Letras da UFMG;

À Capes, pela concessão da bolsa;

E a todas as pessoas que de alguma forma contribuíram para este trabalho.

SUMÁRIO

SUMÁRIO	7
ÍNDICE DE FIGURAS	10
ÍNDICE DE TABELAS.....	11
RESUMO	15
CAPÍTULO 1 INTRODUÇÃO	16
CAPÍTULO 2 CARACTERIZAÇÃO DO FENÔMENO VARIÁVEL.....	19
2.1. Caracterização acústica das variantes	23
2.2. Fenômenos em análise	30
CAPÍTULO 3 QUADRO TEÓRICO	32
CAPÍTULO 4 METODOLOGIA DE PESQUISA.....	39
4.1. A coleta dos dados	39
4.2. A cidade de Itaúna e o bairro das Graças	40
4.3. Descrição da amostra.....	42
4.3.1. Fala.....	42
4.3.2. Escrita.....	45
4.4. Coleta dos Dados de Fala	45
4.4.1. Aspectos técnicos da gravação.....	45
4.4.2. Entrevistas	46
4.4.3. Leitura e testes	46
4.5. Coleta dos Dados de Escrita	48
4.6. Criação dos Bancos de Dados.....	48
4.6.1. Fala.....	48
4.6.1.1. Transcrição dos dados.....	49
4.6.1.2. Seleção das Variantes.....	49
4.6.1.3. Codificações	50
4.6.1.4. Codificação das variantes.....	52
4.6.2. Escrita.....	52
4.6.2.1. Normalização	52
4.6.2.2. Contagem de frequência	53
4.7. Seleção das Variáveis independentes	54
4.7.1 Variáveis sociais	54
4.7.1.1. Gênero e Faixa Etária.....	54
4.7.1.2. Estilo.....	56
4.7.2. Variáveis Internas	56

4.7.2.1.	Contexto seguinte.....	56
4.7.2.2.	Classe da palavra seguinte.....	57
4.7.2.3.	Contexto anterior.....	57
4.7.2.4.	Classe da palavra.....	58
4.7.2.5.	Fricativa alveolar seguinte.....	58
4.7.2.6.	Vogal da variável.....	59
4.7.2.7.	Tonicidade.....	59
4.7.3.	Outras Variáveis.....	59
4.7.3.1.	Frequência.....	60
4.7.3.2.	Indivíduo.....	60
4.7.3.3.	Item lexical.....	60
CAPÍTULO 5 MODELOS DE ANÁLISE QUANTITATIVA.....		62
5.1	O modelo estatístico.....	64
5.2.	Definições preliminares.....	66
5.2.1.	Hipótese nula, nível de significância e p-valor.....	66
5.2.2.	Probabilidade, chance e razão de chances.....	67
5.3.	Modelo de regressão logística.....	69
5.3.1.	Seleção das variáveis independentes.....	71
5.4.	Comparando o Varbrul e o SPSS.....	73
5.4.1.	Diferentes formas de codificação de fatores.....	73
5.4.1.1.	Fator de referência.....	73
5.4.1.2.	Desvio da média.....	75
5.4.2.	Comparando as saídas.....	80
5.4.3.	Interação entre variáveis independentes.....	87
5.4.4.	Avaliação dos softwares.....	89
CAPÍTULO 6 ANÁLISE DOS RESULTADOS.....		92
6.1.	Teste de interações.....	93
6.2.	Interação entre gênero e faixa etária.....	96
6.2.1.	Apagamento da vogal – [1]+vogal ~ [1].....	97
6.2.2.	Velarização da lateral – [1] ~ [ɣ].....	98
6.2.3.	Apagamento da sílaba - [1]+vogal ~ Ø.....	98
6.3.	Seleção do modelo.....	99
6.3.1.	Variável [1]+vogal ~ Ø.....	100
6.3.2.	Variável [1]+vogal ~ [1].....	101
6.3.3.	Variável [1] ~ [ɣ].....	102
6.4.	Variáveis sociais.....	102
6.4.1.	Gênero e faixa etária.....	102
6.4.2.	Estilo.....	106
6.5.	Variáveis lingüísticas.....	108

6.5.1. Contexto seguinte.....	108
6.5.2. Contexto anterior.....	110
6.5.3. Classe da palavra.....	111
6.5.4. Vogal da variável.....	113
6.5.5. Classe da palavra seguinte.....	114
6.5.6. Fricativa alveolar seguinte.....	116
6.5.7. Tonicidade.....	119
6.6. Outras variáveis.....	121
6.6.1. Indivíduo.....	121
6.6.2. Os itens lexicais.....	128
6.6.3. A frequência.....	131
6.7. Teste de avaliação das variantes.....	137
CAPÍTULO 7 CONCLUSÕES FINAIS.....	142
REFERÊNCIAS.....	148
ANEXO 1.....	157
ANEXO 2.....	159
ANEXO 3.....	161
ANEXO 4.....	167
ANEXO 5.....	169
ANEXO 6.....	174
ANEXO 7.....	177
ANEXO 8.....	180
ANEXO 9.....	183
ANEXO 10.....	185
ANEXO 11.....	187
ANEXO 12.....	189
ANEXO 13.....	191
ANEXO 14.....	193
ANEXO 15.....	198
ANEXO 16.....	203
ANEXO 17.....	207
ANEXO 18.....	210

ÍNDICE DE FIGURAS

Figura 1.	Espectrograma: “quando eu to de férias da escola, ou senão <u>ele</u> ” - [e1i]-RH17	24
Figura 2.	Espectrograma: “assaltou só <u>ele</u> mais nada” - [e1]-DM16	25
Figura 3.	Espectrograma: “não sei aonde <u>ele</u> tava indo não ele foi assaltado” - [eɿ]-WH38	25
Figura 4.	Espectrograma: “quando <u>ele</u> fez a curva e viu que o negócio” - [e] - CH33	26
Figura 5.	Espectrograma: “quando eu to de férias da <u>escola</u> , ou senão...” - [iskolə]-RH17	26
Figura 6.	Espectrograma: “onde ce tava, tava na <u>escol</u> ” - [iskol]-AH18.....	27
Figura 7.	Espectrograma: “ligou lá, falou que tinha uma bomba na <u>escola</u> ” - [iskolɿ]-TH18	27
Figura 8.	Espectrograma: “estadual, <u>escola</u> estadual de Itaúna” - [sko]-LM40.....	28
Figura 9.	Mapa de localização da cidade de Itaúna	40
Figura 10.	Qualidade da gravação	46
Figura 11:	CART para a variável dependente [1] ^{+vogal} ~ [1] e variável independente <i>indivíduo</i> , no estilo <i>entrevista</i>	122
Figura 12:	Recorte da Figura 11.....	123
Figura 13:	CART para a variável dependente [1] ^{+vogal} ~ Ø e variável independente <i>indivíduo</i> , no estilo <i>entrevista</i>	125
Figura 14:	CART para a variável dependente [1] ~ [ɿ] e variável independente <i>indivíduo</i> , no estilo <i>entrevista</i>	126
Figura 15:	CART para a variável dependente [1] ^{+vogal} ~ [1] e variável independente <i>indivíduo</i> , no estilo <i>leitura de texto</i>	127
Figura 16:	Gráficos de distribuição dos dados para as variáveis <i>fala</i> , <i>Lael_fala</i> e <i>escrita</i> (gráfico Q-Q Plot, distribuição normal).....	133
Figura 17:	Gráfico de distribuição da frequência dos itens lexicais terminados em sílaba átona final /l/ ^{+vogal} nos dados de fala de Itaúna (Q-Q Plot, distribuição normal).....	134

ÍNDICE DE TABELAS

Tabela 1:	Distribuição da população de Itaúna entre gênero, faixa etária e região.....	41
Tabela 2:	Estratificação da amostra de Itaúna.....	44
Tabela 3:	Cancelamento da lateral intervocálica com e sem presença de /S/ final em Martins (2001).....	59
Tabela 4:	Tabela de Contingência para o gênero na variável [l]+vogal ~ Ø no estilo entrevista.....	67
Tabela 5:	Matriz de desenho para a variável gênero.....	73
Tabela 6:	Matriz de desenho para a variável contexto seguinte	74
Tabela 7:	Tabela de contingência para o contexto seguinte na variável dependente apagamento da sílaba /l/+vogal	75
Tabela 8:	Tabela de contingência para variável dependente apagamento da sílaba /l/+vogal com frequência relativa dos fatores da variável contexto seguinte	78
Tabela 9:	Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ Ø, no estilo entrevista, utilizando o SPSS	81
Tabela 10:	Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ Ø, no estilo entrevista, utilizando o Varbrul	83
Tabela 11:	Comparação entre as saídas do Varbrul e do SPSS para a variável dependente [l]+vogal ~ Ø.....	84
Tabela 12:	Comparação entre as saídas do Varbrul e do SPSS para a variável dependente [l]+vogal ~ [l]	85
Tabela 13:	Comparação entre as saídas do Varbrul e do SPSS para a variável dependente [l] ~ [ɫ]	86
Tabela 14:	Recorte do modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ [l], no estilo entrevista, incluindo o termo da interação entre gênero e faixa etária.....	89
Tabela 15:	Distribuição das variantes nos estilos entrevista, perguntas direcionadas, leitura de texto e leitura de palavras.....	92
Tabela 16:	Distribuição das variantes no estilo <i>entrevista</i>	93
Tabela 17:	Distribuição dos dados para as variáveis dependentes no estilo <i>entrevista</i>	93
Tabela 18:	Tabela de contingência para a variável dependente [l]+vogal ~ Ø com cruzamento das variáveis <i>contexto seguinte</i> e <i>classe da palavra</i>	95
Tabela 19:	Saída do SPSS, com variável dependente [l]+vogal ~ [l] e variáveis independentes gênero e faixa etária.....	97

Tabela 20:	Saída do SPSS, com variável dependente [1]+vogal ~ [1] e variáveis independentes gênero, faixa etária e o termo da interação entre gênero e faixa etária.....	97
Tabela 21:	Saída do SPSS, com variável dependente [1] ~ [1] e variáveis independentes gênero e faixa etária.....	98
Tabela 22:	Saída do SPSS, com variável dependente [1]+vogal ~ \emptyset e variáveis independentes gênero e faixa etária.....	99
Tabela 23:	Saída do SPSS, com variável dependente [1]+vogal ~ \emptyset e variáveis independentes gênero, faixa etária e o termo da interação entre gênero e faixa etária.....	99
Tabela 24:	Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [1]+vogal ~ \emptyset , no estilo entrevista, com termo da interação.....	100
Tabela 25:	Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [1]+vogal ~ [1], no estilo entrevista, com termo da interação.....	101
Tabela 26:	Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [1] ~ [1], no estilo entrevista.....	102
Tabela 27:	Resultados para variável <i>gênero-faixa etária</i> na variável dependente [1]+vogal ~ \emptyset , no estilo entrevista.....	103
Tabela 28:	Resultados para variável <i>gênero-faixa etária</i> na variável dependente [1]+vogal ~ [1], no estilo entrevista.....	104
Tabela 29:	Resultados para variável <i>gênero-faixa etária</i> na variável dependente [1] ~ [1], no estilo entrevista.....	105
Tabela 30:	Tabela de contingência entre a variável <i>estilo</i> e todas as variantes.....	106
Tabela 31:	Tabela de contingência entre os fatores <i>entrevista e leitura de textos</i> , da variável <i>estilo</i> e a variável <i>gênero/faixa etária</i> , para a variável dependente [1]+vogal e [1].....	107
Tabela 32:	Resultados do efeito da variável <i>contexto seguinte</i> na variável dependente [1]+vogal ~ \emptyset	108
Tabela 33:	Resultados do efeito da variável <i>contexto seguinte</i> na variável dependente [1]+vogal ~ \emptyset , excluídos os pronomes.....	109
Tabela 34:	Resultados do efeito da variável <i>contexto seguinte</i> na variável dependente [1]+vogal ~ [1].....	109
Tabela 35:	Resultados do efeito da variável <i>contexto seguinte</i> na variável dependente [1] ~ [1].....	110
Tabela 36:	Cruzamento da variável variável dependente [1] ~ [1] com a variável <i>contexto seguinte</i> , sem pronomes.....	110
Tabela 37:	Resultados do efeito da variável <i>classe da palavra</i> na variável dependente [1]+vogal ~ \emptyset	111
Tabela 38:	Resultados do efeito da variável <i>classe da palavra</i> na variável dependente [1]+vogal ~ \emptyset excluídos os pronomes ele/eles/ela/elas.....	112
Tabela 39:	Resultados do efeito da variável <i>classe da palavra</i> na variável dependente [1] ~ [1].....	112

Tabela 40:	Resultados do efeito da variável <i>vogal da variável</i> na variável dependente [1]+vogal ~ Ø.....	113
Tabela 41:	Distribuição da variável <i>vogal da variável</i> entre pronomes na variável dependente [1]+vogal ~ Ø.....	113
Tabela 42:	Resultados do efeito da variável <i>vogal da variável</i> na variável dependente [1]+vogal ~ [1].....	113
Tabela 43:	Resultados do efeito da variável <i>classe da palavra seguinte</i> na variável dependente [1]+vogal ~ [Ø].....	114
Tabela 44:	Resultados do efeito da variável <i>classe da palavra seguinte</i> na variável dependente [1]+vogal ~ Ø, somente para os itens ele/ela/eles/elas.....	115
Tabela 45:	Resultados do efeito da variável <i>classe da palavra seguinte</i> na variável dependente [1]+vogal ~ [1].....	116
Tabela 46:	Resultados do efeito da variável <i>fricativa alveolar seguinte</i> na variável dependente [1]+vogal ~ Ø.....	116
Tabela 47:	Ocorrências de itens que apresentam /S/ final e apagamento da sílaba final /l/+vogal.....	117
Tabela 48:	Resultados do efeito da variável <i>indeterminação</i> na variável dependente [1]+vogal ~ Ø para o item <i>eles</i>	117
Tabela 49:	Comparação entre itens com /S/ e itens sem /S/ na variável [1]+vogal ~ Ø.....	118
Tabela 50:	Resultados do efeito da variável <i>fricativa alveolar seguinte</i> na variável dependente [1]+vogal ~ [1].....	119
Tabela 51:	Resultados do efeito da variável <i>tonicidade</i> na variável dependente [1]+vogal ~ Ø.....	119
Tabela 52:	Resultados do efeito da variável <i>tonicidade</i> na variável dependente [1] ~ [±].....	120
Tabela 53:	Distribuição das proparoxítonas entre as variantes no estilo entrevista.....	120
Tabela 55:	Resultados do efeito do reagrupamento dos indivíduos na variável dependente [1]+vogal ~ [1].....	124
Tabela 56:	Itens lexicais distribuídos em relação às variantes da sílaba final átona /l/+vogal, no estilo <i>entrevista</i>	128
Tabela 57:	Coeficientes de correlação de Spearman entre as variáveis fala, Lael_fala e escrita.....	134
Tabela 58:	Resultados do efeito da variável <i>freqüência de ocorrência do item</i> no banco de dados de fala de Itaúna na variável dependente <i>probabilidade de não ocorrer</i> [1]+vogal.....	136
Tabela 59:	Resultados do efeito da variável <i>freqüência de ocorrência do item vogal</i> no banco de dados de escrita de Itaúna na variável dependente <i>probabilidade de não ocorrer</i> [1]+vogal.....	136
Tabela 60:	Resultados do efeito da variável <i>freqüência de ocorrência do item</i> no banco de dados do LAEL na variável dependente <i>probabilidade de não ocorrer</i> [1]+vogal.....	136
Tabela 61:	Frases distribuídas de acordo com a avaliação dos falantes.....	139

Tabela 62:	Resultados do efeito da variável <i>avaliação</i> na variável dependente [l]+vogal ~ [l]/[ɫ]/Ø.....	140
Tabela 63:	Resultados do efeito da variável <i>avaliação</i> na variável dependente [l]+vogal ~ Ø.....	140
Tabela 64:	Resultados do efeito da variável <i>avaliação</i> na variável dependente [l]+vogal ~ [l].....	141
Tabela 65:	Resultados do efeito da variável <i>avaliação</i> na variável dependente [l]+vogal ~ [ɫ].....	141

RESUMO

Este trabalho analisa a variação em itens lexicais com sílaba final átona formada por /l/+vogal na região de Itaúna/MG. O objetivo principal é identificar e analisar os fatores que influenciam a variabilidade na fala da comunidade pesquisada. Assumem-se os fundamentos do modelo teórico da fonologia de uso, de base difusionista. Consideram-se também aspectos teórico-metodológicos da sociolinguística variacionista. Os processos variáveis analisados compreendem: o apagamento da sílaba final átona formada por /l/+vogal; o apagamento da vogal da sílaba formada por /l/+vogal e o apagamento da vogal e a velarização da lateral da sílaba formada por /l/+vogal. Para analisar tais processos, foi utilizado um corpus formado por 6.616 ocorrências, coletadas a partir de entrevistas e testes com 16 informantes da comunidade de fala pesquisa. Além desse corpus, foi utilizado um corpus escrito de 2 milhões de palavras, coletadas em 4 jornais da cidade de Itaúna. Como variáveis sociais foram considerados o gênero, a faixa etária e o estilo. Como variáveis internas considerou-se o contexto seguinte, o contexto anterior, a vogal da variável, a tonicidade, a classe do item, a classe do item seguinte e a presença de /S/ na sílaba. Além destas variáveis, consideraram-se também as variáveis: indivíduo, frequência de *token* e o item lexical. A identificação dos efeitos atribuídos às variáveis sociais e linguísticas, bem como influências atribuídas ao indivíduo, permitiu uma análise ampla do que de fato parece estar atuando nos processos variáveis. Apresentaram-se evidências de que características associadas ao uso, como a frequência de *token*, parecem também atuar nos fenômenos de etiologia fonética, além dos fatores internos. Além da análise dos fenômenos variáveis, faz-se uma análise comparativa entre os softwares *Varbrul* e *SPSS*, no capítulo destinado à descrição do modelo estatístico utilizado nesta dissertação.

CAPÍTULO 1

INTRODUÇÃO

Este trabalho tem por objetivo analisar sincronicamente o fenômeno de variação sonora ocorrido em itens lexicais com sílaba final átona formada por lateral seguida de vogal na cidade de Itaúna/MG. Para tal, serão analisadas as ocorrências das seguintes variantes para a seqüência de /l/+vogal final átona:

1. Realização plena da sílaba formada por lateral alveolar seguida de vogal, como em *aquele*
2. Realização da lateral alveolar e não realização da vogal seguinte, como em *aque!*;
3. Velarização da lateral e não realização da vogal seguinte, como em *aquẽ*;
4. Não realização da sílaba formada por lateral + vogal, como em *aquẽ*.

Diversos estudos têm mostrado que o segmento lateral /l/ é bastante sujeito a processos de variação sonora no português do Brasil, tanto na posição de *codá*, conforme Tasca (2002), Vandresen (1999), Espiga (2003) e Cristófaros-Silva e Oliveira (2001), quanto no ataque silábico, conforme Martins (2001) e Correa (1998). Além disso, os fenômenos que atingem o segmento lateral também podem ser identificados em estudos históricos da língua portuguesa, como em Tessyer (1997), Zágari (1988), Williams (1975), Coutinho (2004), Câmara Jr. (1985) e Maia (1986).

Para análise do processo variável ocorrido em Itaúna, conforme caracterizado acima, são assumidos os pressupostos teórico-metodológicos da sociolinguística variacionista, segundo o qual as línguas humanas são inerentemente variáveis e sujeitas a mudanças. Nesta perspectiva, a variação linguística não ocorre de forma aleatória, mas é influenciada por fatores sociais e por fatores linguísticos estatisticamente mensuráveis. Assim, este estudo buscará identificar e analisar que fatores influenciam a variabilidade ocorrida na fala da comunidade pesquisada.

Além dos pressupostos da sociolinguística variacionista, serão considerados os pressupostos da fonologia de uso, especialmente aqueles que tratam da influência da frequência de ocorrência do item na mudança sonora, aspecto normalmente não considerado nos estudos em sociolinguística variacionista.

Esta dissertação organiza-se em 7 capítulos, formados por esta introdução e 6 capítulos resumidamente apresentados nos parágrafos seguintes.

No capítulo 2, caracteriza-se a variável linguística em análise a partir das ocorrências das variantes no corpus coletado em Itaúna. A caracterização das variantes será feita a partir de exemplos reais da fala dos informantes submetidos à análise acústica. A hipótese apresentada neste capítulo é de que a variação ocorrida nos itens terminados em sílaba átona /l/+vogal corresponde a processos variáveis diferenciados.

No capítulo 3 são apresentados e discutidos os pressupostos teóricos considerados nesta pesquisa. São apresentados os fundamentos da sociolinguística variacionista e da fonologia de uso, cada qual com suas especificidades, algumas diferenças e semelhanças e as justificativas para a utilização conjunta de tais modelos.

O capítulo 4 aborda os aspectos metodológicos utilizados nesta pesquisa. São descritas as metodologias de seleção da amostra, de coleta e transcrição dos dados e de codificação das variáveis.

No capítulo 5 expõe-se detalhadamente o modelo estatístico utilizado na análise dos dados coletados, o *modelo de regressão logística*. Propõe-se uma análise dos procedimentos internos utilizados pelo software *Varbrul* para a geração dos pesos relativos e uma comparação dos resultados apresentados por tal software com os resultados de um método convencional de *regressão logística*. Além disso, são analisadas as conseqüências da inclusão e da análise da interação entre variáveis independentes ocorrida no banco de dados. Ao final, apresenta-se uma comparação entre os softwares *Varbrul* e *SPSS*.

No capítulo 6 são apresentados os resultados da análise quantitativa dos dados coletados. Inicialmente, apresenta-se a análise da interação no banco de dados e, em seguida, passa-se à análise de cada variável independente considerada neste estudo tendo em vista a sua influência na variabilidade dos dados de fala coletados. Além das variáveis sociais e linguísticas, são apresentados os resultados da análise do indivíduo e da sua influência nos processos variáveis em análise e os resultados que apontam a influência da

freqüência de ocorrência (*token*) dos itens lexicais nesses processos. No final deste capítulo, apresenta-se a análise do teste de avaliação das variantes, através do qual é possível identificar indícios do grau de estigmatização do processo e também avaliar o grau de consciência social dos fenômenos analisados neste trabalho.

Finalmente, no capítulo 7 são apresentadas as conclusões finais e algumas indicações de pesquisas futuras que possam contribuir numa compreensão mais aprofundada dos fenômenos analisados neste estudo.

CAPÍTULO 2

CARACTERIZAÇÃO DO FENÔMENO VARIÁVEL

As consoantes líquidas no português do Brasil têm sido objeto de diversos estudos, entre eles Oliveira (1983), Tasca (2002), Cristófaros-Silva e Oliveira (2001), Martins (2001), Vandresen (1999), Espiga (2003), entre outros. Neste grupo de sons, encontram-se, no sistema fonológico do português, o tap alveolar (como em caro ['karu]), a lateral palatal (como em alho ['aʎu]) e a lateral alveolar (como em bolo ['bolu]).

Observações do dialeto falado na cidade de Itaúna puderam indicar a existência de um fenômeno lingüístico variável envolvendo itens terminados em sílaba composta de lateral alveolar seguida de vogal, como no caso do pronome *ele*. Tal fenômeno caracteriza-se pela oposição entre a realização da sílaba plena, como em ['eli] e o apagamento da vogal final com velarização do segmento lateral, como em ['eɫ].

A lateral velarizada é um segmento comum em alguns dialetos do português brasileiro. Em Tasca (2002), a autora identifica um percurso histórico da variação do segmento lateral na coda silábica em dialetos do Rio Grande do Sul e caracteriza a velarização como parte de tal percurso. Em tais dialetos, a variável lingüística compõe-se de quatro variantes para o segmento lateral, caracterizadas como *alveolar*, *velar*, *velarizada*, *labializada* e *vocalizada*. No estudo em questão, Tasca identifica o fenômeno como um possível processo de mudança lingüística em progresso.

A análise dos dados coletados em Itaúna indica que nesse dialeto a velarização se restringe à posição final de palavra. Indica também que o processo de velarização parece ser posterior a um processo de apagamento da vogal seguinte, já que foram identificadas ocorrências do tipo ['eɫ], em que havia apagamento da vogal final sem velarização da lateral, mas não foram identificadas ocorrências do tipo ['eɫi]. O processo de apagamento da vogal final, em sílaba formada por /l/+vogal, com /l/ alveolar ou velarizado, também parece ser restrito a sílabas átonas, já que formas tônicas com vogal final apagada também não foram identificadas nos dados. Além disso, parece não haver restrições quanto à classe gramatical no processo de apagamento da vogal, pois ele ocorre em itens de diferentes

classes, como ['faɫ] ou ['fal] para o verbo *falo* e [trã 'kwil] ou [trã 'kwilɫ] para o item *tranquilo*.

Além da sílaba plena, do apagamento da vogal e da variante velarizada, observou-se também que alguns itens apresentavam um apagamento total da sílaba final, manifestando-se, por exemplo, como ê, para *ele*, e aquê, para *aquele*.

A partir da identificação dessas quatro variantes fonéticas no dialeto falado em Itaúna, presentes na sílaba final átona de itens lexicais terminados em /l/+vogal, temos caracterizada uma variável lingüística composta por:

1. Realização plena da sílaba /l/+vogal com lateral alveolar. Ex.: “Quando não era ['eɫi], era o padre (...)” (ele), LM40.
2. Realização do /l/ como alveolar e não-realização da vogal seguinte. Ex.: “Minha mãe nunca foi na [is 'kɔl] por minha causa.” (escola), TH18.
3. Velarização do /l/ e não-realização da vogal seguinte. “Eu acho até que a gente era mais [trã 'kwilɫ].” (tranquilo), RH17.
4. Não realização da sílaba formada por /l/+vogal. “áí foi quando em oitenta e dois que teve [a 'kɛ] virada”. (aquela), EM39.

Outros estudos variacionistas abordaram, de formas diferenciadas, fenômenos envolvendo o segmento lateral. Correa (1998) realizou um estudo sobre o cancelamento do /l/ ou da vogal seguinte a ele em pronomes pessoais de terceira pessoa do caso reto (ele/eles/ela/elas), no português falado na cidade de Belo Horizonte. O autor buscou explicar o fenômeno do cancelamento apoiando-se na hipótese de que ele figura-se como formação de novos clíticos na Língua Portuguesa.

Línguas que apresentam preenchimento obrigatório da posição de sujeito, como o Francês, desenvolveram formas clíticas de sujeito. Estudos sobre o Português Brasileiro têm mostrado uma tendência ao preenchimento desta posição. (Tarallo, 1983, Duarte, 1995). Tais conclusões permitem que se formule a hipótese de que o fenômeno em estudo aqui possa estar associado a essa tendência, configurando uma nova estratégia de preenchimento da posição de sujeito. (CORREA, 1998, p. 3)

Nesse estudo, considerarei todos os itens lexicais terminados em /l/+vogal em posição final átona. Dessa maneira, a análise do fenômeno não se restringirá, assim como

não se restringe nos dados reais, a itens pronominais de 3ª pessoa do caso reto. Ao contrário, a variação será analisada tendo em vista todas as possibilidades de sua ocorrência no recorte desta pesquisa.

Martins (2001) estudou o cancelamento de consoantes líquidas (/l/ e /r/) em posição intervocálica na cidade de Belo Horizonte. Além de um estudo que possa fornecer dados empíricos de uma região diferente da pesquisada pela autora, nosso estudo diferencia-se por propor um estudo específico de itens com sílaba final formada por lateral mais vogal. Em Martins (op.cit.), tais itens são analisados juntamente com outros que possuem a variável no interior da palavra. Diferentemente, em Itaúna, a forma velarizada, em especial, parece ter ocorrência exclusiva nas sílabas finais átonas. Em Martins (op.cit.), a forma velarizada não foi encontrada, o que pode indicar que tal variante não está presente na fala de Belo Horizonte.

Os estudos de Tasca (2002), Vandresen (1999) e Espiga (2003) tratam da variação da lateral na coda silábica no estado do Rio Grande do Sul. Em comum, esses estudos discutem o fenômeno da vocalização da lateral velarizada no dialeto sulista identificando o “status” da mudança, se está em progresso ou se trata-se de variação estável nas diversas regiões do estado. Tasca (2002) aponta quatro variantes para o segmento lateral na coda silábica que, segundo ela, configuram-se como quatro etapas da evolução histórica do segmento: uma lateral alveolar, uma lateral velar, uma lateral velarizada-labializada e, finalmente, um glide posterior. Espiga (2003), num estudo sobre o português de fronteira do Rio Grande do Sul, amplia a variável analisada em Tasca (op.cit.) inserindo-lhe uma variável em que a lateral não se realizava, caracterizando assim um zero fonético. Dessa forma, tem-se em Espiga (op.cit.) uma variável com as seguintes variantes:

$$/ l / \rightarrow [l] \sim [ɫ] \sim [ɫ^w] \sim [w] \sim \emptyset^1$$

Na região urbana de Itaúna onde foram feitas as gravações para esta pesquisa, não há variação em itens lexicais como aqueles pesquisados por Tasca (2002), Espiga (2003) e Vandresen (1999), como mal, sal, celta, sul. Nestes itens, parece haver hoje um fenômeno categórico em que a lateral transformou-se em um glide posterior, assim como descrito em Oliveira (1983), num estudo sobre a variação nas líquidas na cidade de Belo Horizonte;

¹ Espiga (2003:p. 255)

Cristófaros-Silva e Oliveira (2001), num estudo sobre a vocalização da lateral no português do Brasil e Quednau (1993)² apud Tasca (2002), num estudo sobre o segmento lateral na cidade de Porto Alegre.

Se considerarmos a realização das variantes em que a vogal não é realizada, teríamos a lateral na coda silábica e um processo que poderá assemelhar-se aos processos descritos nos estudos sobre o dialeto do Rio Grande do Sul.

Andrade (1997) afirma que no dialeto da cidade de Lisboa ocorre uma velarização da lateral no ataque silábico, como no item *ele*. No processo descrito em Andrade, diferentemente do que ocorre em Itaúna, o segmento vocálico seguinte se mantém.

Vários autores já constataram os percursos históricos do segmento lateral na língua portuguesa. Tessyer (1997) e Zágari (1988) afirmam que o cancelamento da lateral alveolar intervocálica na passagem do latim para o português ocorreu em diversos itens, como nos exemplos: *dolor* → *dor*; *salire* → *sair*; *avolo* → *avô*; *malo* → *mau*. Diferente do que ocorre em Itaúna, nos casos citados, o cancelamento não se restringe à posição átona.

De acordo com Williams (1975) e Coutinho (2004), os pronomes pessoais de terceira pessoa do singular do português contemporâneo originaram-se do latim clássico. As formas *ele* e *ela* vieram dos demonstrativos latinos *ille* e *illa*, respectivamente. De acordo com esses autores, as formas *el*, indicando *ele*, e *eis*, indicando *eles*, já eram encontradas no português arcaico. Câmara Jr (1985, p. 98) afirma que nos pronomes “o plural com *-s* é uma criação portuguesa pelo padrão do plural com os nomes”. Da mesma forma, nos pronomes demonstrativos de terceira pessoa já se encontravam, no português arcaico, as formas *aquel* e *aques*, referindo-se respectivamente a *aquele* e *aqueles*. Com relação às formas nominais, Câmara Jr. (1985) descreve algumas passagens do latim popular para o português em que houve perda do segmento lateral, entre eles destacamos: *caelu* > *céu* e *pūlu* > *pó*. Williams (1975) afirma que o “l intervocálico do latim clássico caiu em português: *dolōrem* > *door* > *dor*; *salire* > *sair*; *sōles* > *sóis*” Em Coutinho (2004), o autor apresenta formas do português originadas do latim em que houve queda do segmento vocálico posterior à lateral em posição átona final, como *generale* > *geral* e *hospitale* > *hospital*.

² QUEDNAU, Laura Rosane. *A lateral pós-vocálica no português gaúcho: análise variacionista e representação não-linear*. Porto Alegre: 1993.

De acordo com Maia (1986, p. 498), a variante velarizada está presente no português “desde a fase mais antiga da história da língua” e é proveniente do latim, que já possuía uma consoante lateral velarizada. A vocalização da lateral, segundo a autora, provém de uma realização velar da lateral já no século XVI. Maia (op.cit.) descreve ainda processos históricos da queda da lateral intervocálica e considera tal processo como um traço característico do galego-português, presente em textos do século XIII.

A partir das referências históricas e dos estudos envolvendo a variação na lateral, podemos dizer que a variação ocorrida na cidade de Itaúna nos itens terminados em sílaba átona final formada por /l/+vogal é um fenômeno recorrente na língua portuguesa.

2.1. Caracterização acústica das variantes

A análise acústica dos segmentos laterais que compõem a variável lingüística analisada neste estudo tem por objetivo, principalmente, fornecer parâmetros acústicos para a identificação das formas variantes nos dados coletados na cidade de Itaúna. Assim, interessa-nos caracterizar e diferenciar acusticamente o segmento lateral alveolar seguido de vogal; o segmento lateral sem a presença da vogal; a velarização da lateral sem a presença da vogal e a não realização da sílaba final formada por /l/+vogal.

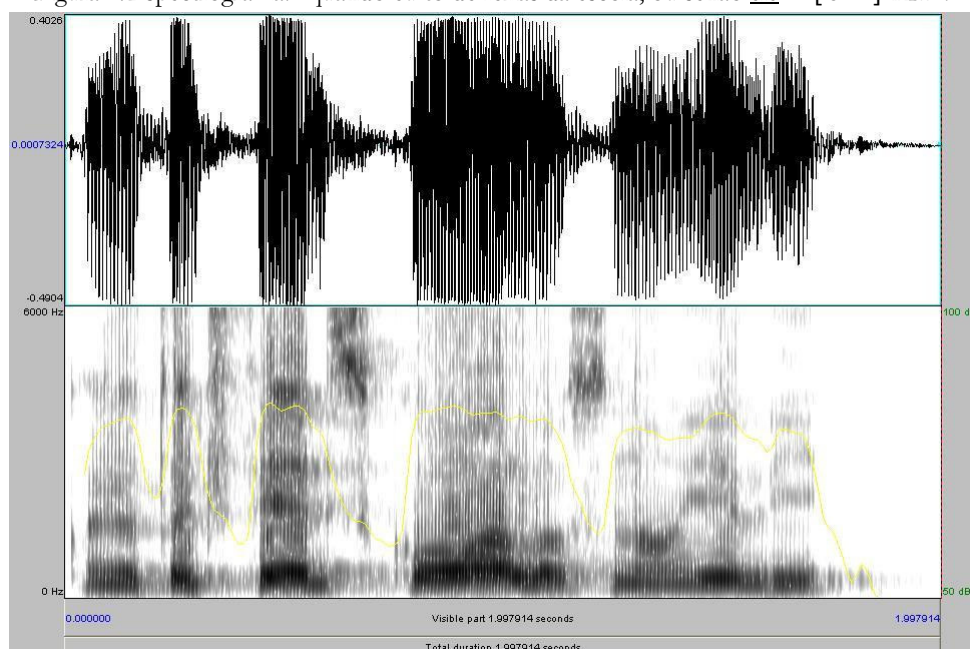
Em Ladefoged e Maddieson (1996), os autores afirmam que o segmento lateral é caracterizado pela presença de formantes bem definidos; o primeiro formante apresenta frequências muito baixas e o segundo formante apresenta frequências mais centrais, variando de acordo com o ponto de oclusão ou com a posição da língua. O terceiro formante apresenta amplitude forte e frequências altas. Assim, podemos dizer que a diferenciação entre as laterais é dada, principalmente, pela altura do segundo formante. Além disso, os autores afirmam que, na presença de uma vogal adjacente, uma mudança abrupta pode ser observada nos formantes devido à presença de coarticulação.

Espiga (2003, p. 258) caracteriza o segmento lateral velarizado em relação à lateral alveolar. Segundo o autor, na velarização,

- ✓ F2 é mais baixo do que em [l]
- ✓ F1 é mais alto do que em [l]
- ✓ A diferença F2 – F1 é menor do que em [l]

Tendo como aparato as informações em Ladefoged e Maddieson (1996) e em Espiga (2003), acredito ser possível diferenciar acusticamente as variantes estudadas na cidade de Itaúna. Para isso, serão analisados alguns espectrogramas da fala natural, coletada por meio de entrevistas. Analisarei as ocorrências das variantes nos itens *ele* e *escola*. Os espectrogramas apresentados foram gerados pelo software Praat, versão 4.4.06³. Os recortes feitos nas entrevistas e a redução do ruído foram feitos com o auxílio do software Cool Edit 2000 v.1.1. As amostras analisadas têm duração de 2 segundos e frequências entre 20Hz e 8000Hz.

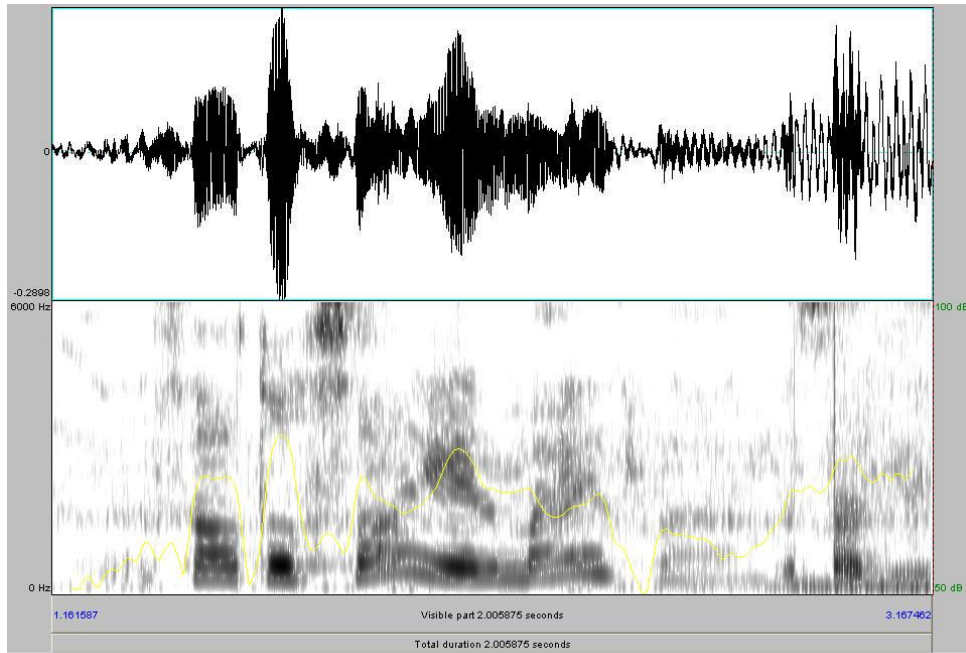
Figura 1. Espectrograma: “quando eu to de férias da escola, ou senão ele” - [eli]-RH17



e l i

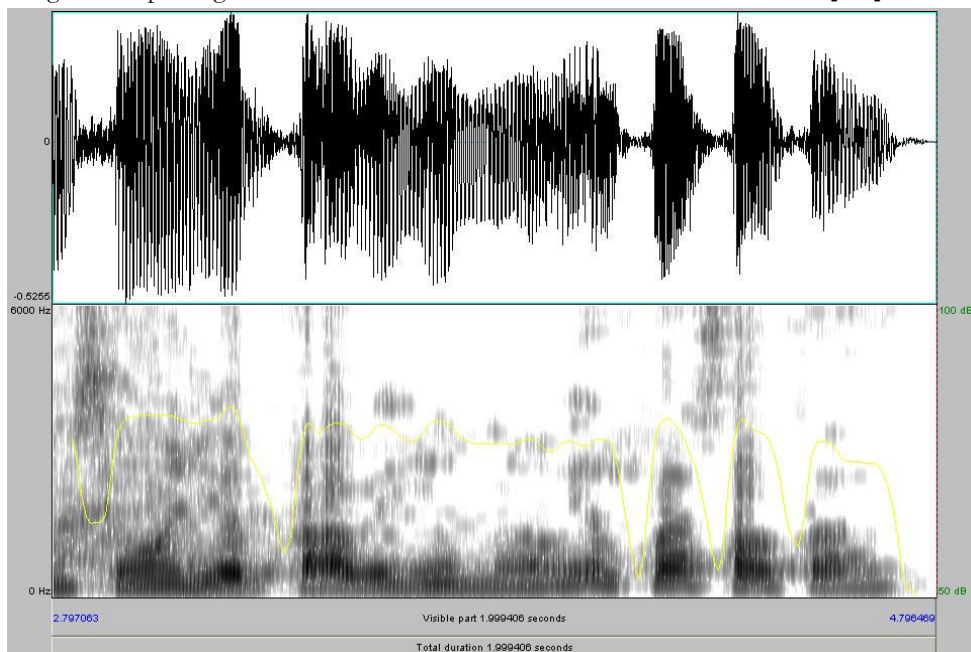
³ Disponível gratuitamente em <http://www.fon.hum.uva.nl/praat/>

Figura 2. Espectrograma: “assaltou só ele mais nada” - [eɪ]-DM16



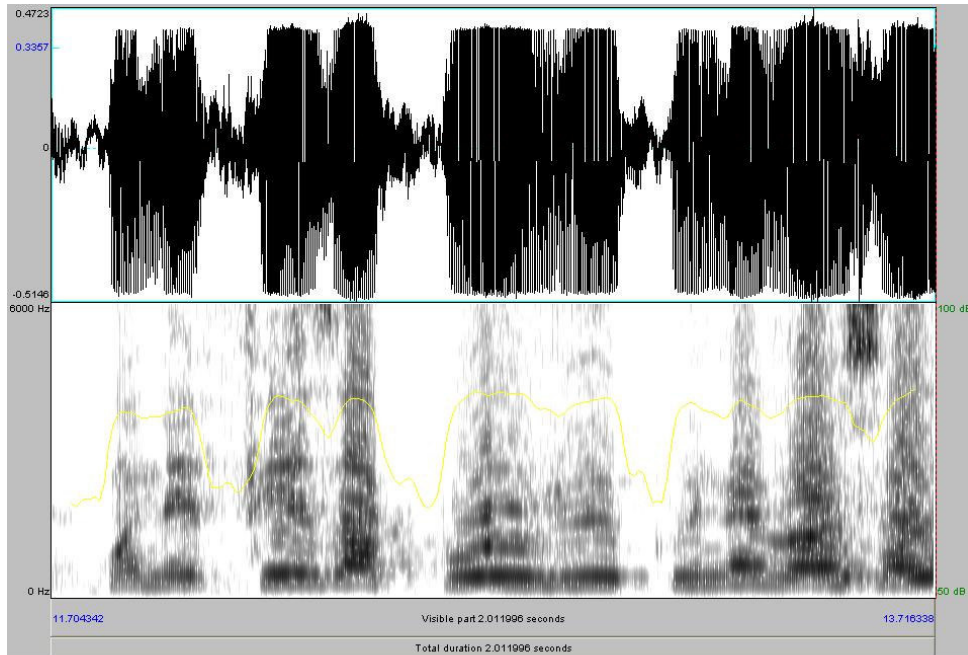
eɪ

Figura 3. Espectrograma: “não sei aonde ele tava indo não ele foi assaltado” - [eɪ]-WH38



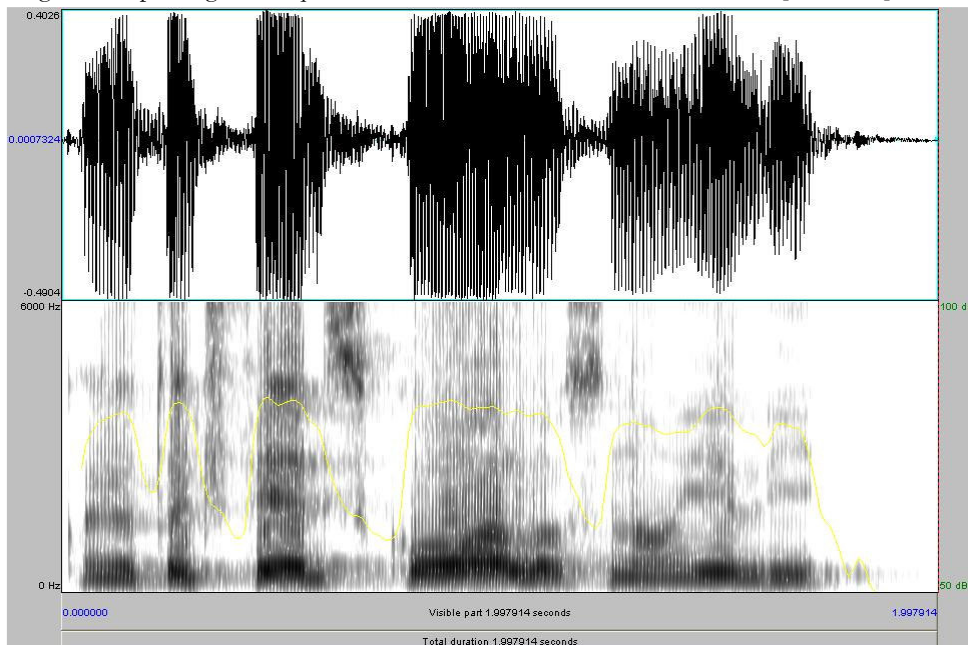
eɪ

Figura 4. Espectrograma: “quando ele fez a curva e viu que o negócio” - [e] - CH33



e

Figura 5. Espectrograma: “quando eu to de férias da escola, ou senão...” - [iskɔlə]-RH17



i s k ɔ l ə

Figura 6. Espectrograma: “onde ce tava, tava na escola” - [iskol]-AH18

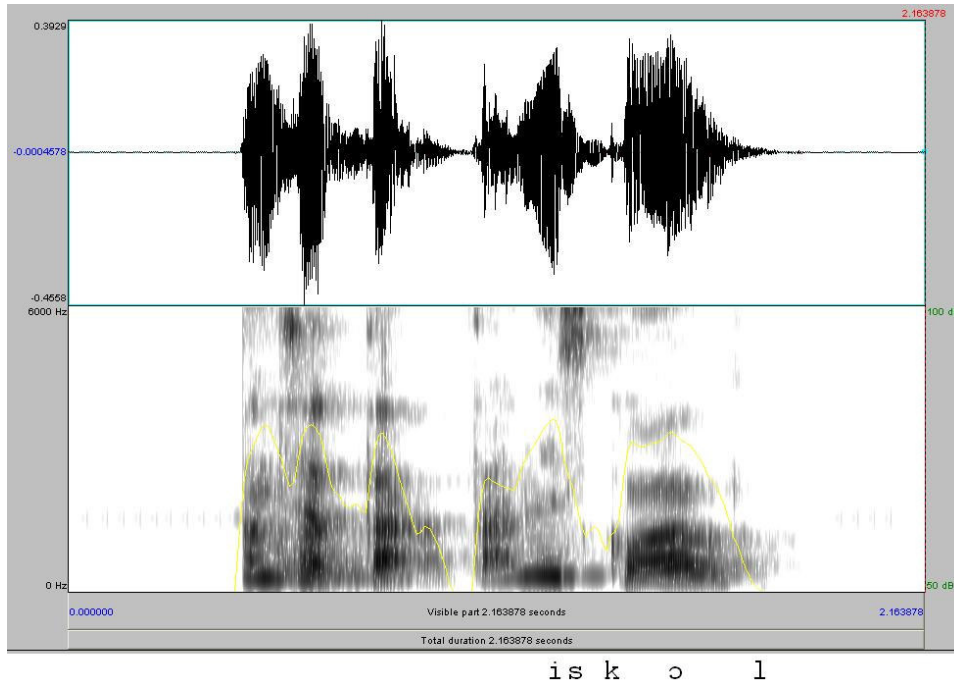


Figura 7. Espectrograma: “ligou lá, falou que tinha uma bomba na escola” - [iskol]-TH18

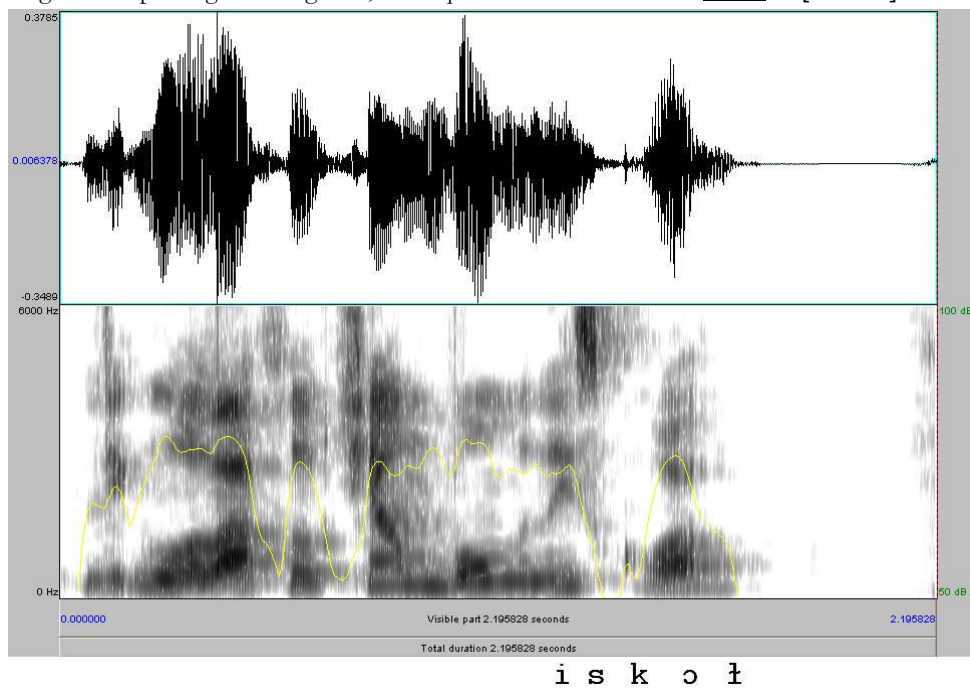
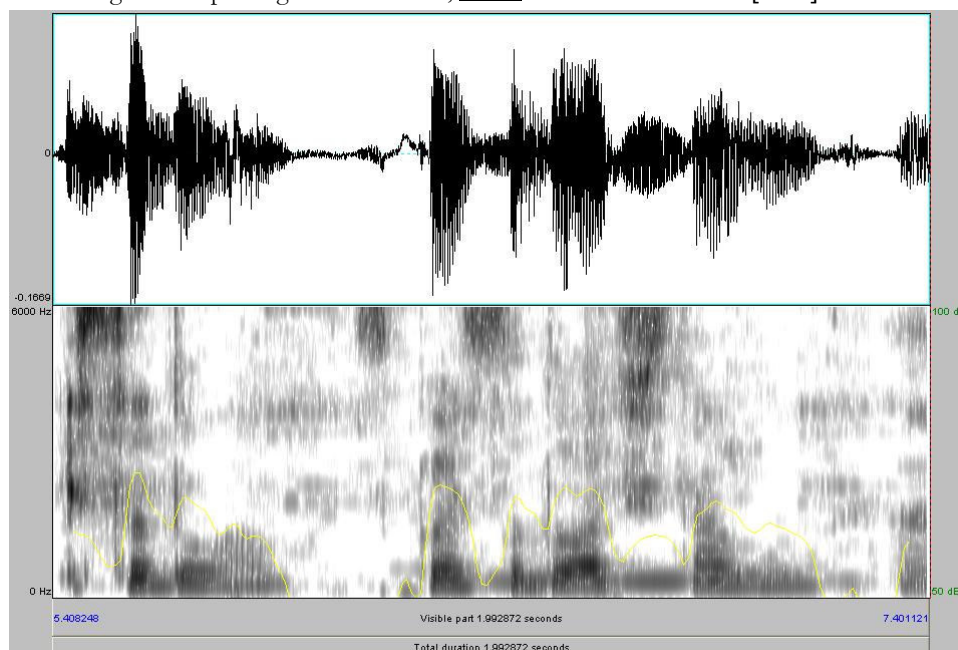


Figura 8. Espectrograma: “estadual, escola estadual de Itaúna” - [skɔ]-LM40

s k ɔ

As figuras 1 a 4 referem-se às variantes de /l/+vogal no item lexical *ele*. Na figura 1 temos a realização plena da sílaba formada por lateral+vogal, com a presença de formantes bem definidos para o segmento lateral, bem como para a vogal seguinte [i]. Na figura 2, vemos os formantes da lateral, mas os formantes da vogal seguinte não são produzidos. Neste exemplo, os formantes posteriores aos da lateral correspondem ao início da consoante [m] seguinte. As médias das freqüências da lateral na figura 2 são F1=502Hz; F2=1.171Hz e F3=2.728Hz. Na figura 3 pode-se perceber visualmente a ausência de formantes para o segmento vocálico [i] e um abaixamento da freqüência do segundo formante no segmento lateral, o que está de acordo com as características da velarização descritas em Tasca (2003). As médias das freqüências dos formantes neste exemplo são F1: 512Hz ; F2: 962Hz e F3: 2.667Hz. Na figura 4, os formantes da lateral não estão presentes.

As figuras 5 a 8 referem-se às variantes de /l/+vogal no item lexical *escola*. No espectrograma da figura 5 vemos com clareza os formantes da lateral seguidos dos formantes da vogal [ə] seguinte. Já na figura 6 apresentam-se os formantes do segmento lateral, mas não há formantes para o segmento [ə]. No exemplo seguinte, figura 7, assim como na figura 3, pode-se perceber um abaixamento na freqüência do segundo formante da lateral em relação à lateral nas figuras 6 e 5, o que caracteriza a presença de um segmento velarizado. Na figura 8, vêem-se os formantes da vogal [ɔ], mas não há formantes para os segmentos que compõem a sílaba final /l/+vogal como na figura 5.

Os exemplos analisados indicam que é possível categorizar acusticamente as variantes analisadas neste trabalho. A categorização de tais variantes nos dados coletados em Itaúna foi feita principalmente a partir da análise dos seus espectros e tendo como parâmetros as características apresentadas nos exemplos acima. Em alguns casos raros, em que havia dúvida na análise do sinal acústico, consideraram-se também aspectos perceptivos, com identificação das variantes a partir da produção repetida do item analisado em testes informais de percepção.

Além das variantes analisadas acima, notei a presença de alguns sons que não se enquadravam exatamente na categorização das variantes. Encontrei, por exemplo, casos em que o segmento lateral era praticamente imperceptível no espectrograma mas, do ponto de vista perceptivo, parecia haver a presença de uma lateral. Em tais ocorrências, categorizei a ocorrência como tendo presente a lateral e ausente a vogal seguinte. Em outros casos, notei a presença de um ditongo, como no caso de [da 'kɛə], para o item *daquela*, provavelmente fruto do cancelamento da líquida intervocálica, fenômeno já analisado em Martins (2001) e Correa (1998) e recorrente no português, como descrito em Maia (1986). Tais ocorrências foram categorizadas como não-realização da sílaba final formada por /l/+vogal. Encontrei também casos em que a vogal final da sílaba /l/+vogal apresenta-se com duração muito reduzida, mas nos quais os formantes da vogal ainda podiam ser identificados. Tais ocorrências foram categorizadas como realização plena da sílaba final.

A dificuldade de categorização, dada pela existência de observações que não se encaixam acusticamente nas variantes, nos dá fortes indícios de um processo foneticamente gradual presente no segmento em análise.

Na frase “ela lembra quando ela era” (IM15), a presença de um som vocálico seguinte à sílaba final formada por /l/+vogal, como em “ela era”, cria um ambiente propício à não realização da vogal seguinte ao /l/, como ocorrido no exemplo acima, [ɛ 'lɛrə] (ela era). Tal ocorrência foi categorizada como não realização da vogal seguinte ao /l/. Da mesma forma, a presença de uma lateral seguinte pode favorecer a não-realização da sílaba, como também ocorrido na frase acima, [ɛ 'lɛbrə] (ela lembra). Tal ocorrência foi categorizada como não realização da sílaba final formada por /l/+vogal. A opção por não desconsiderar tais dados se deu porque, da mesma forma, foram encontrados diversos dados em que tais ambientes ocorriam e que, no entanto, a sílaba final ocorria de forma plena, como “na passagem do ano, ela estava” (LM17) e “ia prender

ele lá” (EM39), ambos os itens grifados pronunciados plenamente. A possibilidade de interferência do segmento seguinte na realização das variantes será investigada detalhadamente no capítulo 6.

2.2. Fenômenos em análise

Uma questão interessante a ser respondida neste trabalho refere-se à caracterização do fenômeno ou dos fenômenos ocorridos nos itens terminados em /l/+vogal na cidade de Itaúna. Para discutir tal questão, proponho as seguintes perguntas:

1. As variantes encontradas em Itaúna fazem parte de um processo único de variação sonora, no qual teríamos somente uma variável linguística, composta de quatro variantes como as caracterizadas acusticamente na seção anterior?

ou

2. As variantes encontradas em Itaúna fazem parte de processos independentes e com especificidades próprias?

Minha hipótese inicial é de que as variantes encontradas fazem parte de processos distintos de variação sonora, caracterizados por especificidades que possibilitam assim classificá-los.

Por hipótese, o percurso de tais processos está sintetizado nos esquemas abaixo para o item lexical *aquela*:

variante 1 variante 2 variante 3

1. aquela: [a'kɛlɐ] → [a'kɛl] → [a'kɛɿ]

variante 1 variante 4

2. aquela: [a'kɛlɐ] → [a'kɛə] → [a'kɛ]

No processo 1, ocorre um apagamento da vogal seguinte ao /l/ e, posteriormente, uma velarização da lateral. No processo 2, ocorre um cancelamento da líquida intervocálica e, posteriormente, um processo de apagamento da vogal final. O processo 2 apresenta uma redução maior, visto o número de segmentos em relação ao processo 1.

Diante dos processos identificados acima, proponho a análise simultânea de três variáveis dependentes, descritas abaixo:

Variável 1 → [l]+vogal ~ Ø (variantes 1 e 4): Essa variável será chamada de *apagamento da sílaba final*. A análise da variante em que a lateral é cancelada, mas a vogal é mantida, não será feita porque tais variantes ocorreram raríssimas vezes no corpus.

Variável 2 → [l]+vogal ~ [l] (variantes 1 e 2): Essa variável será chamada de *apagamento da vogal*.

Variável 3 → [l] ~ [ɭ] (variantes 2 e 3): Essa variável será chamada de *velarização da lateral*. Ela incluirá todos os dados em que a vogal final não ocorre e em que a lateral ocorre variavelmente como alveolar ou como velarizada.

Mesmo na hipótese da existência de processos com especificidades próprias, o fato de tais processos ocorrerem nos mesmos itens lexicais e em contextos lingüísticos idênticos leva-nos a acreditar que tais processos estabelecem fortes relações. Assim, a análise das variáveis dependentes precisa ser feita concomitantemente, para que possamos ter uma idéia da relação estabelecida entre os processos. Farei isso, na apresentação dos resultados, mostrando a realização das variantes nas variáveis dependentes, dada a presença dos fatores das variáveis independentes selecionadas neste estudo. O reflexo dessa organização do trabalho será a possibilidade de uma visão mais abrangente de como cada variável independente interfere na realização das variantes.

CAPÍTULO 3

QUADRO TEÓRICO

Neste trabalho trataremos de pressupostos teórico-metodológicos da sociolinguística variacionista, desenvolvidos a partir da década de 60 por William Labov. Labov desenvolveu o modelo de descrição e interpretação da variabilidade linguística, o que possibilitou a explicação de tal variabilidade, colocando-a como inerente ao componente linguístico. Além da sociolinguística variacionista, abordaremos também as propostas dos modelos de *linguagem em uso*, especialmente a *fonologia de uso*, desenvolvida por Joan Bybee, que considera que uso e gramática interagem e que a frequência de uso atua nos processos de mudança. Considerarei as propostas da fonologia de uso para uma melhor explicação de processos em termos da frequência de uso de itens lexicais dentre outros aspectos.

A sociolinguística variacionista concebe a língua como um sistema heterogêneo e a variação e a mudança linguística como inerentes a esse sistema (Labov, 1972, p. 223). Além disso, considera que o aparente caos gerado pela variação é influenciado por fatores internos ao sistema e por fatores de natureza social. Assim, o estudo da variação linguística, nessa perspectiva, objetiva determinar que fatores são os responsáveis pela variação e pela mudança numa determinada língua e que fatores sociais são responsáveis pela propagação da mudança.

Em Labov (2002), o autor propõe que a variação linguística é favorecida por fatores internos e implementada socialmente.

The triggering event that leads to extensive systematic change is the insertion or removal of a category from a sub-system in a direction determined by unidirectional constraints on linguistic change. The principle of maximal dispersion then applies within that sub-system as a driving force for continued change. This may apply to semantic, syntactic, morphological or phonological systems, though our knowledge of the principles governing phonological change is perhaps the most highly developed.

(...) The major acceleration of diffusion within a community is the result of the reinterpretation of social category differences as gender differences. Once a social category correlation is reinterpreted as a gender difference, it is subject to the basic mechanism of reinterpretation as the unmarked community level, leading to rising levels of linguistic change for the community as a whole. (LABOV, 2002).

Uma análise variacionista permite identificar a relação entre a produção de uma variante associada probabilisticamente a algum fator lingüístico ou social. A variação é, a priori, um processo motivado por um conjunto de fatores identificáveis e mensuráveis estatisticamente e, portanto, não é aleatório. A variável sociolingüística é, por sua vez, dependente de outras variáveis e é, por isso, chamada variável dependente. As variáveis que influenciam a variável dependente são chamadas de variáveis independentes. Para mensurar a influência das variáveis independentes na variável dependente, grande parte dos estudos variacionistas tem utilizado um método estatístico chamado de regressão logística, que possibilita investigar a mudança na variável dependente correspondente à mudança nas variáveis independentes. O modelo estatístico utilizado na análise variacionista será discutido no capítulo 5.

O objeto de estudo da análise variacionista é a comunidade de fala. Labov (1972) afirma que:

The speech community is not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms; these norms may be observed in overt types of evaluative behavior, and by the uniformity of abstract patterns of variation which are invariant in respect to particular levels of usage. (LABOV, 1972, p. 120-121).

Apesar do objeto de estudo da sociolingüística variacionista não ser especialmente a fala do indivíduo, mas da comunidade de fala, o estudo dos aspectos lingüísticos da fala individual são importantes, entre outras coisas, na medida em que possibilitam a identificação dos líderes e dos inovadores nos processos de mudança, ou seja, aqueles responsáveis por levarem e por iniciarem um processo de mudança na comunidade.

De acordo com Guy e Zilles (texto não publicado),

A hipótese básica da sociolingüística, baseada no conceito de comunidade de fala, é de que os membros de uma comunidade compartilham essencialmente a mesma gramática, inclusive os efeitos dos contextos lingüísticos sobre um processo variável.

Assim, as variáveis independentes lingüísticas não exerceriam efeitos diferenciados por classe social, gênero ou faixa etária. Este assunto será retomado na discussão dos resultados referentes ao indivíduo. Sankoff (1988, p.992) afirma que existe uma pressuposição implícita de que não existe interação entre os fatores associados aos falantes e os outros fatores que representam aspectos da estrutura lingüística. Ainda que um

processo de mudança esteja mais avançado em um determinado grupo do que em outro, em uma comunidade de fala socialmente estratificada, se um fator lingüístico for favorecedor de tal processo, ele o será em todos os grupos que tenham desencadeado o processo e na mesma direção, mas pode ser que tal processo não tenha atingido determinado grupo social ainda.

Uma importante questão desenvolvida pela sociolingüística variacionista refere-se à implementação social da mudança. Labov (1972) apresenta dois tipos de mudanças: mudanças *from below* (abaixo do nível de consciência social do fenômeno) e mudanças *from above* (acima do nível de consciência social do fenômeno). Em Labov (1994), tais mudanças são assim caracterizadas:

Mudanças *from above*

are introduced by the dominant social class, often with full public awareness. Normally, they represent borrowing from other speech communities that have higher prestige in the view of the dominant class. Such borrowings do not immediately affect the vernacular patterns of the dominant classes or other social classes, but appear primarily in careful speech, reflecting a superposed dialect learned after the vernacular is acquired. Frequently the newly borrowed linguistic features are inconsistent with the vernacular system, and their use involves correlated changes in other features. (LABOV, 1994, p.78)

Mudanças *from below*

changes from below are systematic changes that appear first in the vernacular, and represent the operation of internal, linguistic factors. At the outset, and through most of their development, they are completely below the level of social awareness. No one notices them or talks about them, and even phonetically trained observers may be quite unconscious of them, and even phonetically trained observers may be quite unconscious of them for many years. It is only when the change are nearing completion that members of the community become aware of them. Changes from below may be introduced by any social class, although no cases have been recorded in which the highest-status social group acts as the innovating group. (LABOV, 1994, p.78)

Labov (op.cit, p. 542) estabelece que as mudanças *from below* são mudanças sonoras regulares, resultantes "de uma transformação gradual de uma característica fonética simples de um fonema em um espaço fonético contínuo"⁴. Labov afirma ainda que as mudanças *from above* são mudanças que se dão por difusão lexical e são resultantes "da substituição abrupta de um fonema por outro em itens que contenham esse fonema". Tais mudanças seriam características de estágios finais e poderiam apresentar condicionamento lexical ou

⁴ Em Kiparsky (1995), o autor propõe uma discussão sobre esta caracterização que não será analisada aqui.

gramatical. Além disso, caracterizam-se pelo alto grau de consciência social do fenômeno.

A base da discussão, levantada em Labov (1994), encontra-se numa discussão mais antiga entre duas correntes do pensamento lingüístico: a corrente neogramática e a corrente difusionista, comentadas a seguir. Em Labov (1994), o autor apresenta uma tentativa de conciliação de tais correntes.

Na perspectiva desenvolvida pelos neogramáticos, a mudança lingüística atinge todo o léxico abruptamente e não permite exceções. A mudança sonora ocorreria de forma foneticamente gradual, a partir de mudanças graduais na estrutura sonora dos segmentos. Nessa perspectiva, o léxico não exerce influência na mudança, pois as mudanças ocorrem no som. Se um som muda, ele deve mudar em todos os itens sem seleção lexical. Os problemas encontrados pelos neogramáticos, ocasionados por itens que não satisfaziam os postulados propostos, são explicados em termos de analogia ou de empréstimos de outras línguas ou permanecem sem explicação.

Na perspectiva difusionista, o léxico exerce um papel fundamental na mudança. Propõe-se que a propagação da mudança, ao contrário do que dizem os neogramáticos, se dê de forma lexicalmente gradual, para todo tipo de mudança, atingindo o léxico item a item. Na fonologia de uso, modelo de base difusionista, a mudança ocorre no som em determinada palavra.

Oliveira (1995) esclarece que ambos os modelos consideram os efeitos fonéticos e os efeitos lexicais. Segundo o autor, a diferença entre eles reside "na ordenação relativa dos efeitos lexical e fonético" (op.cit., p. 78). Sobre esta questão Oliveira afirma que:

(...) é interessante observar que os lingüistas de inclinação neogramática resistam à idéia do léxico como controlador primário de uma mudança sonora, mas não à idéia de que, uma vez disparada, uma vez disparada a mudança possa ser implementada lexicalmente. Do mesmo modo, nenhum trabalho de inclinação difusionista dirá que o contexto e/ou o efeito fonético não deva(m) ser levado(s) em conta; o que não se aceita é que eles sejam uma explicação para uma determinada mudança sonora. (OLIVEIRA, 1995, p. 78).

A proposta de Labov (1994) busca responder aos indícios de ambas as perspectivas ao estabelecer a abrangência de cada uma na mudança sonora. Assim, para Labov, as mudanças *from below* ocorreriam da forma como propõem os neogramáticos e as mudanças *from above* ocorreriam da forma como propõem os difusionistas.

Diferentemente do que propõe Labov, Phillips (1984) e Bybee (2001) afirmam que as mudanças *from below* também ocorrem por difusão lexical. Da mesma forma, Oliveira (1991, p. 103) afirma que "(...) all sound changes are lexically implemented, that is, there are no neogrammarian sound changes".

Diversos trabalhos sobre o português do Brasil, Oliveira (1992), Viegas (2001), Cristófaros-Silva (2001), entre outros, têm corroborado a hipótese de que todas mudanças sonoras ocorrem por difusão lexical, e não somente as mudanças *from above*, como propõe Labov (1994) e Labov (2001). Viegas (2001, p. 34) afirma que, ao contrário do que ocorre numa perspectiva neogramática, "no modelo difusionista, as 'exceções' não são vistas como um problema pois, a seleção da mudança sendo lexical, espera-se que os itens todos não tenham o mesmo comportamento".

Bybee (2002) afirma que, além de serem sempre lexicalmente graduais, as mudanças também podem ser foneticamente graduais:

The view of lexical diffusion espoused by both Wang and Labov assumes that a change that diffuses gradually through the lexicon must be phonetically abrupt. This is a necessary assumption if one is to accept a synchronic phonological theory that has phonemic underlying representations. Words can change one by one only if the change is a substitution of phonemes in such theory. The discovery that sound change can be both phonetically gradual and lexically gradual forces different view of the mental representation of the phonology of words" (BYBEE, 2002, p. 271).

O modelo proposto em Bybee (2001), a fonologia de uso, compartilha com a sociolinguística variacionista a noção de variação como inerente à natureza das línguas e que gramáticas não são unidades estáticas, já que estão em constantes processos de mudança. Entretanto, a noção de mudança na perspectiva da fonologia de uso difere-se da noção adotada em Labov (1994). Em Labov (1994, p. 542), como já mencionado acima, mudança sonora é definida como "o resultado de uma transformação gradual de uma característica fonética simples de um fonema em um espaço fonético contínuo", ou seja, uma mudança sonora ocorre no som, o que faz com que todos os itens lexicais que contenham determinado som sejam atingidos pela mudança. Na fonologia de uso, diferentemente, uma mudança não atinge necessariamente todos itens, ela se dá no som em determinadas palavras. Segundo Bybee (2003),

Sound change that are complete can be identified as regular or not, depending upon whether or not they affected all lexical items existing at the time of the change. Ongoing changes cannot be designated as regular or not, since they are

not complete. However, one can reference the typical characteristics of a change to project whether it will be regular or not. (BYBEE, 2003, p. 63).

Na fonologia de uso, os caminhos pelos quais as mudanças irão passar são determinados pela linguagem em uso, pela frequência no uso de formas e padrões sonoros e pelos esquemas generalizantes. Bybee (2001) assume os pressupostos de outro modelo teórico, chamado Modelo de Exemplos. Para ela, o detalhe fonético faz parte da representação mental e tal representação é altamente redundante. Nesse modelo a unidade de representação é a palavra ou grupo de palavras. As variantes que compõem uma variável lingüística são armazenadas redundantemente no léxico e estabelecem fortes relações sonoras e semânticas. Para a fonologia de uso, o estudo da linguagem passa pela relação entre a sua substância (som e significado), seu uso e pela formação de esquemas generalizantes. Dessa forma, atribui-se ao uso, e à frequência de uso em especial, um papel fundamental na mudança lingüística.

Bybee (2001) afirma que a frequência interfere de forma direta na mudança lingüística e destaca que há a atuação da frequência de ocorrência do item (*token*) e da frequência de ocorrência de padrões sonoros (*type*). A frequência de *type* contribui para a produtividade do padrão: padrões mais frequentes parecem ser mais produtivos do que padrões menos frequentes. Juntamente com a frequência de *type*, a frequência de *token* também ocupa um lugar fundamental na fonologia de uso. Segundo Bybee (2001),

In particular, the frequency with which individual words or sequences of word are used and the frequency with which certain patterns recur in a language affects the nature of mental representation and in some cases the actual phonetic shape of words (BYBEE, 2001, p. 1)

A frequência de *token* refere-se à frequência de um item particular no total do *corpus*. Cada ocorrência de um item lexical particular em um *corpus* reforça a sua representação. O item *ele*, por exemplo, no corpus escrito de Itaúna, ocorreu 3.263 vezes em 2 milhões de palavras. Se localizarmos outro item, *vale*, por exemplo, temos que sua ocorrência foi de 376 vezes em dois milhões. A observação dos itens e suas respectivas ocorrências no corpus analisado possibilita classificá-los, comparativamente, como itens de alta, média ou baixa frequência. Não há um consenso em relação à posição onde se deva fazer o corte relacionado às frequências dos itens. Em Bybee (2001, p. 41) a autora apresenta um estudo em que itens com frequência de 149 e 492 ocorrências por milhão são definidos como itens de alta frequência, itens com frequência de 21, 51 e 91 ocorrências por milhão, como itens de média frequência e itens com frequência de 0 e 4 ocorrências por milhão, como itens de

baixa frequência. Na página 112, a autora apresenta um estudo onde se define o corte em 35 ocorrências por milhão. Abaixo de 35 correspondem aos itens de baixa frequência, acima de 35 correspondem aos itens de alta frequência. Discutirei os critérios utilizados para os cortes nesta dissertação na seção relacionada à análise dos resultados de frequência.

Bybee (2001) afirma que, em mudanças de etiologia fonética, os primeiros itens a serem atingidos são aqueles de frequência de *token* mais alta. Assim, a mudança sonora nos itens analisados em Itaúna, caso ela se caracterize como de etiologia fonética, deve estar ocorrendo, ou deve ter ocorrido, primeiramente em itens de frequência mais alta. Bybee (2001, p. 65-69) aponta três critérios para a identificação de mudanças de etiologia fonética: (1) os graus de variabilidade entre as alternâncias: de acordo com a autora, processos de etiologia fonética tendem a apresentar variantes com diferenças fonéticas menores, o processo de mudança de um som para o outro é gradual; (2) o status das exceções: as exceções para processos fonéticos ocorrem em itens de baixa frequência ou em itens identificadas como empréstimos ou como altamente formais e (3) o domínio de aplicação: processos foneticamente motivados geralmente aplicam-se através da fronteira de palavras desde que haja ambiente favorecedor, enquanto que outros processos restringem-se ao nível da palavra.

Segundo Phillips (2001), *apud* Bybee (2001), mudanças gramaticais e baseadas em analogia comportam-se diferentemente de mudanças fonéticas, com relação à frequência. Nos processos baseados em analogia, a alta frequência cria itens com "alto nível de força lexical" e que, por isso, assumem uma independência em relação aos processos de mudança e, assim, dificilmente mudam. A explicação para esse efeito encontra-se no argumento de que a consolidação da representação mental do item é mais robusta devido à frequência com que tal item é acessado.

Como pôde ser visto neste capítulo, há diferenças entre os modelos *sociolinguística variacionista* e *fonologia de uso*, como por exemplo, em relação à concepção de mudança e à interferência da frequência de *token* e de *type* nos processos de variação e mudança linguística. Assumirei aqui os fundamentos do modelo teórico da *fonologia de uso*.

CAPÍTULO 4

METODOLOGIA DE PESQUISA

4.1. A coleta dos dados

Nesta pesquisa foram adotadas duas metodologias para coleta dos dados, ambas submetidas à análise quantitativa; uma composta de dados de fala, coletada a partir de uma amostra socialmente estratificada; e outra a partir de dados de escrita, coletada a partir de textos extraídos de alguns veículos de comunicação escrita da cidade de Itaúna. A opção por métodos quantitativos está em consonância com modelos de análise lingüística que considerem a língua em uso e a variação lingüística encontrada em dados reais como objetos de estudo importantes para os estudos lingüísticos. Tanto na sociolingüística variacionista quanto nos modelos de língua em uso, o estudo da variação lingüística se dá por meio da observação e análise de como os fenômenos lingüísticos realmente ocorrem, em situações reais de fala. A pesquisa quantitativa possibilita uma visão de um universo a partir de um subconjunto representativo da população como um todo. A partir da pesquisa quantitativa é possível medir, com certa confiabilidade, que influência outros fatores exercem sobre o fenômeno que está sendo estudado. Nos estudos de variação lingüística, a análise quantitativa contribui para avaliar se os fatores, selecionados previamente, segundo hipóteses de trabalho do pesquisador, interferem na realização de uma determinada variante em detrimento de outra.

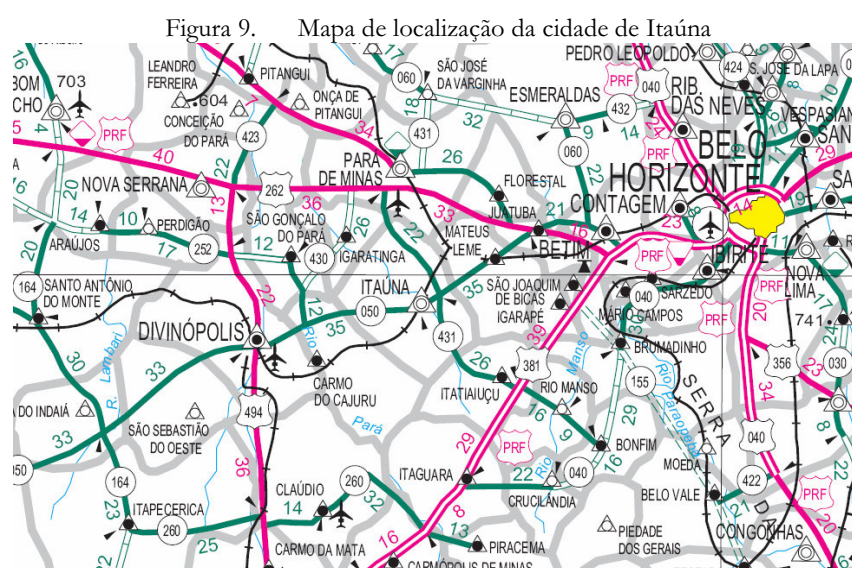
Para este trabalho, a composição da amostra que forneceu os dados de fala foi feita de forma controlada, a partir de características sociais previamente definidas. A opção por uma amostra estratificada se deve ao fato de que, dentro do universo pesquisado, uma seleção totalmente aleatória poderia aumentar a chance de enviesamento dos resultados. Assim, foi feita uma distribuição equitativa dos informantes. As pesquisas em sociolingüística variacionista têm mostrado que a implementação de uma variável lingüística segue, inclusive, padrões socialmente determinados, seja pela idade, pela classe social, pelo grau de escolaridade, pelo gênero, ou por qualquer outro fator de natureza social. Diante disso, torna-se importante fazer um controle amostral que considere um número exaustivo de estratos sociais, já que eles podem exercer influência sobre a implementação do fenômeno em análise e podem contribuir no enviesamento dos resultados, caso eles não

sejam considerados e controlados. A formação dos estratos deve ser feita de forma que todos os indivíduos estejam neles alocados e que não haja nenhum indivíduo em dois estratos diferentes dentro de um mesmo grupo de fatores.

A composição da amostra para fornecimento de dados de escrita se deu a partir de alguns veículos de comunicação escrita da cidade de Itaúna. A partir dessa amostra objetivava-se obter um número significativo de itens lexicais para que se pudesse considerar os efeitos de frequência sobre os processos variáveis em análise. Optou-se por essa coleta para que os dados pudessem refletir de forma mais concreta um universo dos itens mais utilizados na cidade, mesmo que na escrita, e que se pudesse obter desses itens uma contagem de frequência de suas ocorrências. Além dos dados de escrita coletados, utilizarei também o banco de dados de fala do Lael/PUC-SP que, apesar de tratar-se de uma região diferente, também poderá contribuir para que sejam feitas comparações com o corpus de fala coletado em Itaúna. As ressalvas para a utilização de um corpus de fala de outra região serão tratadas em seções subsequentes.

4.2. A cidade de Itaúna e o bairro das Graças

Itaúna está localizada na região oeste de Minas Gerais, a 82km de Belo Horizonte. Limita-se com os municípios de Pará-de-Minas, Igaratinga, Itatiaiuçu, Mateus Leme e Carmo do Cajuru, conforme pode ser visto nos mapas abaixo:



Fonte: DNIT - Departamento Nacional de Infra-Estrutura de Transportes. *Mapa Rodoviário de Minas Gerais*, 2002.

Com relação aos aspectos históricos do povoamento da cidade, segundo Nogueira (2003), a região começou a ser povoada no início do século XVIII à beira do Rio São João. Na década de 1750, já contava com aproximadamente 100 habitantes, entre portugueses, seus descendentes e escravos. Em 1765 foi construída a primeira igreja, cuja padroeira, Senhora de Santana, deu o primeiro nome ao lugar, Povoação Nova de Santana do São João Acima. Só no início do século XX, em 1901, o povoado foi emancipado da cidade vizinha, Pará de Minas, e em 1915 foi elevada à categoria de cidade.

De acordo com dados do censo demográfico realizado pelo IBGE, em 2000, a cidade possuía 76.862 habitantes, sendo 38.138 homens e 38.724 mulheres. A população urbana corresponde a 93,38% do total da população. A tabela 1 apresenta a distribuição da população de Itaúna por gênero, faixa etária e região:

Tabela 1: Distribuição da população de Itaúna entre gênero, faixa etária e região

Faixa Etária	Região Urbana		Região Rural		Total
	Homens	Mulheres	Homens	Mulheres	
0 a 4	2.961	2.948	228	218	6.355
5 a 9	3.244	3.107	247	239	6.837
10 a 14	3.294	3.230	262	212	6.998
15 a 19	3.641	3.337	260	256	7.494
20 a 24	3.384	3.283	207	200	7.074
25 a 29	2.888	2.929	223	178	6.218
30 a 34	2.982	3.021	215	191	6.409
35 a 39	3.024	3.037	230	202	6.493
40 a 44	2.598	2.677	191	142	5.608
45 a 49	1.993	2.136	141	107	4.377
50 a 54	1.596	1.702	135	108	3.541
55 a 59	1.139	1.298	103	81	2.621
60 a 64	849	1.108	98	77	2.132
65 a 69	747	927	77	55	1.806
70 a 74	502	721	56	45	1.324
75 a 79	320	447	37	26	830
80 ou mais	238	462	28	17	745
Total	35.400	36.370	2.738	2.354	76.862

Fonte: IBGE: disponível em <http://www.ibge.gov.br/>

Dados da Fundação João Pinheiro⁵ indicam que o índice de desenvolvimento humano da cidade coloca-a como a décima segunda cidade do estado de Minas Gerais com relação à qualidade de vida da população.

Os informantes entrevistados são moradores do Bairro das Graças. Tal bairro pertence à região central da cidade e possui 2.731 habitantes, o que corresponde a 3,5% da

⁵ Disponíveis em <http://www.fjp.gov.br/>

população total do município. A escolha da cidade de Itaúna foi em parte pessoal, por ter morado nesta cidade grande parte da minha infância, e em parte porque as observações iniciais que motivaram tal estudo estavam relacionadas à velarização da lateral, fenômeno não encontrado em outros locais, como em Belo Horizonte, onde já se havia feito estudos sobre a realização do segmento lateral. O bairro das Graças foi escolhido pela facilidade no contato com os informantes, já que eu conhecia moradores deste bairro e, principalmente, por ser ele um dos bairros mais antigos de Itaúna, o que permitiria que lá fossem encontrados adultos nascido no bairro.

4.3. Descrição da amostra

4.3.1. Fala

Foi considerada uma amostra socialmente estratificada em dois níveis: gênero e faixa etária. Marconi e Lakatos (1996) afirmam que mesmo uma amostra socialmente estratificada deveria obedecer a uma seleção aleatória dos indivíduos dentro dos limites dos estratos, para que houvesse representatividade e significância das escolhas em relação à população e para que a amostra não fosse viciada, ou seja, que não apresentasse interferências do pesquisador em sua seleção. A seleção dos informantes foi realizada de forma bastante controlada, para que os resultados pudessem refletir o mais fielmente possível o comportamento lingüístico do grupo analisado.

A seleção dos informantes foi feita a partir de indicações de uma única pessoa, que foi intermediadora entre pesquisador e pesquisados. A partir das pessoas contactadas, selecionou-se o grupo de onde os dados seriam coletados. A seleção desse grupo se deu tendo em vista um conjunto pré-determinado de características, conforme descrito a seguir:

Fatores considerados:

- ✓ Faixa Etária: foram selecionados informantes de duas faixas etárias assim distribuídas: 8 jovens – entre 15 e 20 anos – e 8 adultos – entre 30 e 40 anos;
- ✓ Gênero: foram selecionadas 8 mulheres e 8 homens;

Fatores Não-Considerados:

A seleção dos informantes deveria ser feita de modo que as variáveis sociais não pesquisadas fossem controladas. Dessa forma, seriam selecionados informantes de apenas um estrato dos grupos de fatores não pesquisados, conforme descrito a seguir:

- ✓ Grupo Social: todos os entrevistados deveriam pertencer ao mesmo grupo social (este fator foi avaliado em termos da condição de vida dos informantes);
- ✓ Escolaridade: todos os entrevistados deveriam ter o segundo grau completo ou segundo grau em curso, no caso dos mais jovens;
- ✓ Região: todos os entrevistados deveriam ser nascidos no Bairro das Graças e nunca terem morado em outra cidade (desejava-se que não tivessem morado em outro bairro);
- ✓ Procedência dos pais: desejava-se que os entrevistados tivessem pais nascidos na cidade de Itaúna.

Segue abaixo a tabela com todos os informantes considerados nessa pesquisa.

Tabela 2: Estratificação da amostra de Itaúna

Informante	Faixa Etária	Gênero	Grupo Social	Profissão	Escolaridade	Morou em outra cidade	Morou em outro bairro	Procedência dos pais
BM40	Adultos	Feminino	Médio	Professora primária	2º grau completo	Não	Não	Itaúna
LM40			Médio	Dona de casa	2º grau completo	Não	Não	Itaúna e Varginha
DM35			Médio	Dona de Casa	2º grau completo	Não	Não	Igaratinga e Itaúna
EM39			Médio	Colunista de Jornal	2º grau completo	Não	Não	Itaúna e Mateus Leme
AH34		Masculino	Médio	Alfaiate	2º grau completo	Não	Não	Itaúna
RH39			Médio	Cabeleireiro	2º grau completo	Não	Não	Itaúna
WH38			Médio	Desempregado	2º grau completo	Não	Não	Itaúna
CH33			Médio	Aposentado	2º grau completo	Não	Não	Itaúna e Pará-de-Minas
NM16	Jovens	Feminino	Médio	Estudante	2º grau incompleto	Não	Não	Itaúna e Belo Horizonte
IM15			Médio	Estudante	2º grau incompleto	Não	Não	Itaúna
DM16			Médio	Recepcionista	2º grau incompleto	Não	Não	Itaúna
LM17			Médio	Professora de violão	2º grau incompleto	Não	Não	Itaúna e Bom Despacho
FH16		Masculino	Médio	Estudante	2º grau incompleto	Não	Não	Itaúna e Itaguara
RH17			Médio	Estudante	2º grau incompleto	Não	Não	Divinópolis e Igaratinga
TH18			Médio	Técnico de Informática	2º grau completo	Não	Não	Itaúna e Carmópolis
AH18			Médio	Estudante	2º grau incompleto	Não	Não	Itaúna

Todas as entrevistas foram feitas com o consentimento livre e esclarecido dos informantes, conforme estabelece a resolução 196/96 do Conselho Nacional de Saúde. A eles foi informada a natureza da estudo e de que a utilização da entrevista seria exclusivamente para fins de pesquisa. Também foi garantida que a entrevista seria utilizada de forma confidencial e que eles poderiam recusar-se a participar da entrevista ou retirar

seu consentimento a qualquer momento sem nenhuma penalização. No Anexo 1, apresentamos uma cópia do *Termo de Consentimento Livre e Esclarecido*.

4.3.2. Escrita

A amostra para coleta dos dados de escrita foi composta de edições de quatro jornais de Itaúna, de periodicidade semanal: Brexó, Fanzine, Tribuna e S'passo. Foram utilizadas 110 edições do Jornal Brexó, 40 edições do Jornal Tribuna e todo o conteúdo disponível nos *sites* dos Jornais Fanzine e S'passo. Do total do corpus, 60% corresponde ao Jornal Brexó, 21% corresponde ao Jornal Fanzine, 16% corresponde ao Jornal Tribuna e 3% corresponde ao Jornal S'passo. Todos os dados, inclusive os disponíveis na Internet, foram coletados com a autorização prévia dos proprietários dos jornais. Os quatro jornais selecionados apresentam características semelhantes quanto ao conteúdo, contendo cadernos de coluna social, notícias da política local e nacional, polícia, saúde, classificados, veículos, esportes e agenda cultural. O jornal Fanzine apresenta diferenças em termos de conteúdo, como entrevistas mais longas com personalidades, notícias relacionadas à ufologia e conteúdos relacionados às artes, como música, teatro e cinema.

4.4. Coleta dos Dados de Fala

4.4.1. Aspectos técnicos da gravação

Para a gravação das entrevistas foi utilizado um microfone omnidirecional de lapela, modelo Lesson ML-70. As entrevistas foram gravadas em formato de áudio digital (wav) com o gravador modelo Mega Stick I da MSI. Esse gravador apresenta taxas de amostragem entre 8000Hz e 48000Hz e resolução de 4bits. A taxa de amostragem reflete a quantidade de amostras por segundo. De acordo com Zurmely (2001);

(...) a quantidade de amostras por unidade de tempo de um sinal, chamada taxa e/ou frequência de amostragem, deve ser maior que o dobro da maior frequência contida no sinal a ser amostrado, para que possa ser reproduzido integralmente sem erro de aliasing. A metade da frequência de amostragem é chamada frequência de Nyquist e corresponde ao limite máximo de frequência do sinal que pode ser reproduzido. (ZURMELY, 2001).

A voz humana, conforme Fischer (2000), apresenta frequências entre 40Hz e 8000Hz. Dessa forma, a frequência de amostragem ideal para digitalização da voz humana deve ser maior do que 16000Hz. Esse valor permite que todos os formantes sejam captados na gravação e, conseqüentemente, que se possa obter bons espectrogramas para análise acústica, caso o nível de ruído do ambiente não seja muito alto.

A taxa de amostragem utilizada na coleta de dados foi de 44100Hz, conforme a figura abaixo, que contém informações dos arquivos gerados pelo gravador digital:

Figura 10. Qualidade da gravação

Propriedade	Valor
Áudio	
Taxa de bits	177kbps
Tamanho da amostra de áudio	4 bit
Canais	1 (mono)
Taxa de amostragem de áudio	44 kHz
Formato de áudio	IMA ADPCM

4.4.2. Entrevistas

As entrevistas foram gravadas na residência de cada um dos informantes, posteriormente a um contato inicial em que foi marcado um horário para que a entrevista pudesse ser feita. Nos contatos com os informantes, a palavra “entrevista” não foi utilizada, em vez disso, utilizei-se o termo “conversa”, que teria como assunto a cidade de Itaúna e algumas histórias conhecidas pela pessoa entrevistada. O roteiro da entrevista (Anexo 2) apresenta questões relacionadas à cidade, à história de vida do informante, de seus amigos e familiares, à infância, à escola, ao aumento da criminalidade na cidade, entre outros. As informações prestadas pelo informante serviram de motivo para que lhe fosse perguntado sobre uma experiência pessoal dele em relação àquele assunto. Acredito que a forma como as entrevistas foram feitas e as questões tratadas nela possibilitaram uma coleta bastante significativa e próxima de uma fala cotidiana do entrevistado.

4.4.3. Leitura e testes

Após cada entrevista, os informante foram submetidos a uma seqüência de leituras e de testes (Anexos 3). Aplicados na ordem abaixo:

1) Perguntas direcionadas

Informava-se ao entrevistado que ele seria submetido a uma espécie de jogo em que seria dada uma definição de uma palavra. O informante deveria dizer do que se tratava. O objetivo era obter dos informantes um conjunto de itens lexicais pré-determinados, terminados em /l/+vogal, em um estilo de fala mais formal do que o estilo obtido na entrevista e menos formal do que os estilos de leitura. A lista de questões encontra-se no Anexo 3.

2) Leitura de Texto

Antecipadamente, produziram-se dois textos que continham diversos itens terminados em [l]+vogal. Procurei inserir itens de várias classes de palavras, com tonicidade e ambiente fonético seguinte diversificado. Foi pedido ao informante que lesse os textos da forma mais natural possível e em voz alta, para que pudesse ser gravado. Os textos encontram-se no Anexo 3.

3) Leitura de palavras

Nessa etapa foram mostrados diversos cartazes, cada um contendo uma única palavra terminada em /l/+vogal. As palavras foram lidas pelos informantes em seqüência, à medida que o cartaz correspondente era mostrado. A lista de palavras encontra-se no Anexo 3.

d) Teste de avaliação

Foram gravadas diversas frases previamente (Anexo 3). Cada frase foi repetida duas vezes, em posições diferenciadas e não subseqüentes. Em cada uma das vezes alterava-se somente a produção dos itens terminadas em /l/+vogal, tendo em vista as formas variantes estudadas. Tomou-se o cuidado de tentar produzir as frases da forma mais natural possível, mantendo, inclusive, características dialetais em itens não pesquisados, como “mininu” (menino), “passarin” (passarinho) e “ingravidá”, (engravidar).

Inicialmente, fazia-se uma pequena exposição ao informante que objetivava mostrar a ele a possibilidade de, às vezes, conseguir identificar o grau de escolaridade de uma pessoa pelo jeito que ela fala. Os informantes, utilizando um fone de ouvidos, ouviam cada uma das frases isoladamente e avaliavam a fala em relação ao grau de instrução do falante. Nesse ponto eles deveriam escolher entre duas opções: muito, se ele atribuísse aquela fala a

alguém que tivesse estudado muito; e pouco, se ele atribuísse aquela fala a alguém que tivesse estudado pouco.

O objetivo desse teste era avaliar a sensibilidade dos informantes à variável pesquisada e, ao mesmo tempo, tentar identificar que variantes poderiam ser consideradas estigmatizadas.

4.5. Coleta dos Dados de Escrita

Os textos escritos foram coletados a partir de edições de quatro jornais locais. Os dados de dois jornais foram coletados diretamente de arquivos digitais da versão enviada para as bancas. Os dados dos outros dois jornais foram coletados das páginas publicadas na Internet, que continham a última publicação enviada para as bancas e algumas publicações anteriores.

4.6. Criação dos Bancos de Dados

Os banco de dados foram criados a partir do áudio digital coletado nas entrevistas e a partir dos textos de jornais. As especificações do computador utilizado na criação desses bancos de dados foram: processador AMD Sempron 3000, memória 512MB e placa de som 16bits. Utilizou-se também um teclado multimídia com tecla play/pause e um fone de ouvido.

4.6.1. Fala

O banco de dados de fala foi criado a partir de uma seqüência de etapas. Após os dados terem sido gravados, em formato digital, e transferidos para o computador, todas as entrevistas foram transcritas e codificadas segundo o conjunto de variáveis independentes pré-definidas e as variantes lingüísticas utilizadas pelo falante. Seguem abaixo os passos para criação do banco de dados de fala.

4.6.1.1. **Transcrição dos dados**

A transcrição de dados foi feita seguindo os padrões ortográficos da língua portuguesa. As variantes, tanto do item analisado, quanto dos demais itens, não foram transcritas da forma como o falante realmente as havia produzido, mas de acordo com a ortografia padrão. Mantive as demais características da fala de cada informante, como concordâncias verbais, nominais e regências, exatamente da forma como o informante as havia produzido. Uma frase produzida como: “Tem os poblema del, oh gente, todo mundo tem poblema. Ês já fez muita coisa que eu desagradei.” (em39) foi transcrita dessa forma: “Tem os problema dele, oh gente, todo mundo tem problema. Eles já fez muita coisa que eu desagradei.”. Optou-se por essa forma de transcrição por dois motivos principais: 1) ela possibilitaria uma contagem automatizada de frequência dos itens falados nas entrevistas considerando como iguais ocorrências de itens como “poblema” e “problema” e 2) possibilitaria uma localização automatizada das formas terminados em /l/+vogal.

As transcrições foram feitas no editor de textos TextPad. Adotamos como padrão os símbolos “ / ” para as falas do entrevistado e “ – “ para as falas do entrevistador. A atribuição desses símbolos foi arbitrária e justifica-se somente para simplificar a referência ao falante nas mudanças de turno.

4.6.1.2. **Seleção das Variantes**

Objetivou-se, com uma transcrição ortográfica das entrevistas, possibilitar que os itens a serem analisados fossem encontrados de forma automatizada. Uma busca automática pode ser feita com base em padrões gerais comuns aos itens em análise. No nosso caso, interessava-nos encontrar os itens terminados em /l/+vogal, seguidos ou não de /s/ e somente na fala do informante. A partir de uma seqüência de comandos feita no Microsoft Word, possibilitou-se uma busca automática dos itens a serem analisados com uma taxa relativamente baixa de erros. Tais erros referiam-se a itens que não seriam analisados (como ali e zulus) e puderam ser corrigidos no momento em que foram inseridas as codificações e a classificação das variantes. Ao final do processo de seleção dos itens, tinha-se um arquivo final de cada informante, em formato de tabela, em que, em cada linha, a última palavra do campo terminava em /l/+vogal, com ou sem /s/, e tinha sido falada pelo informante. Tal procedimento, além de diminuir o tempo gasto para a localização dos itens, otimizou a busca dos mesmos e evitou erros na sua localização.

4.6.1.3. Codificações

Para que a análise estatística pudesse ser feita, cada ocorrência, de cada uma das variantes em análise deveria ser codificada em relação a cada uma das variáveis independentes selecionadas. As codificações foram feitas no Microsoft Excel, cada coluna representava uma variável e cada linha representava um dos itens terminados em /l/ +vogal produzidos pelos informantes. As codificações para a variável dependente e para as variáveis independentes, justificadas nas seções subseqüentes, foram feitas da seguinte forma:

1. Formas variantes

- 0 - [l]+vogal → forma plena
- 1 - [l] → apagamento da vogal
- 2 - [ɫ] → apagamento da vogal e velarização da lateral
- 3 - ∅ → apagamento da sílaba

2. Gênero

- 0 - masculino
- 1 - feminino

3. Idade

- 0 - jovens
- 1 - Adultos

4. Informante

- 1 - ah18
- 2 - bm40
- 3 - ah34
- 4 - ch33
- 5 - dm16
- 6 - fh16
- 7 - lm17
- 8 - lm40
- 9 - dm35
- 10 - em39
- 11 - im15
- 12 - nm16
- 13 - rh17
- 14 - rh39
- 15 - th18
- 16 - rh38

5. Estilo
 - 1 – entrevista
 - 2 – perguntas direcionadas
 - 3 – leitura de texto
 - 4 – leitura de palavras

6. Contexto Seguinte
 - 1 – Consoante
 - 2 – Vogal
 - 3 – Pausa

7. Contexto Anterior
 - 0 – vogal alta
 - 1 – vogal baixa
 - 2 – vogal média

8. Classe da Palavra
 - 0 – nome
 - 1 – pronome
 - 2 – verbo

9. Classe da Palavra Seguinte
 - 0 – verbo auxiliar
 - 1 – verbo não-auxiliar
 - 2 – nome
 - 3 – pausa

10. Vogal na variável
 - 0 – alta posterior
 - 1 – alta anterior
 - 2 – baixa

11. Tonicidade
 - 0 – paroxítona
 - 1 – proparoxítona

12. Presença de /S/ final
 - 0 – ausente
 - 1 – presente

Além das variáveis independentes já citadas, também foi criada uma coluna com códigos para os itens lexicais. Cada item recebeu um número, de 1 a 149, enumerados em ordem alfabética, conforme apresentado no Anexo 4. Isso foi feito para que os itens pudessem ser observados isoladamente.

4.6.1.4. Codificação das variantes

As variantes foram classificadas a partir da análise acústica de cada uma de suas ocorrências e, em alguns casos, a partir de critérios perceptivos. A análise acústica foi feita tendo como base os espectrogramas produzidos pelo software Cool Edit 2000. O Cool Edit 2000 não é um software construído especificamente para análise acústica, ao contrário, sua aplicação está voltada principalmente à produção e edição de arquivos de áudio. Entretanto, apesar dele ser restrito para análise acústica em relação a outros softwares construídos para esta finalidade, pois ele não exibe informações como intensidade e frequência dos formantes, o Cool Edit possui algumas facilidades e recursos que justificaram seu uso, já que a única exigência nesta fase da pesquisa era de que o software produzisse bons espectrogramas. O software permite a redução de ruídos de forma bastante simples e eficaz. Ruídos causados pelo ambiente no momento em que foram feitas as gravações ou produzidos pelo próprio gravador são facilmente retirados do arquivo de áudio. Outro benefício foi a facilidade no manuseio de arquivos grandes e de longa duração, que permitiu analisar os arquivos de áudio da forma e no tamanho que eles foram gravados.

4.6.2. Escrita

4.6.2.1. Normalização

Os dados de escrita provenientes dos jornais *Brexó*, *Fanzine*, *S'passo* e *Tribuna* tiveram que passar por alguns processos de normalização para que pudessem ser lidos pelo software de contagem de frequências de palavras (*Frequência NumaBoa*). Os textos conseguidos com cada um dos jornais apresentavam-se em formatos diferenciados: html, pdf e p65. Dessa forma, a primeira providência foi converter todos os formatos em um formato único, de texto simples, para que pudessem ser colocados em um arquivo único e para que pudessem ser lidos pelo *Frequência NumaBoa*. Para a conversão dos formatos originais em txt foram utilizados os seguintes softwares: *Adobe PageMaker 7*, para conversão dos arquivos p65 em pdf; *Advanced PDF2TXT*, para conversão dos arquivos pdf em txt; e *Html2Txt*, para conversão dos arquivos html em txt.

Posteriormente, pelos problemas causados pelas conversões e por problemas de formatação dos textos originais, o arquivo txt que incluía todos os dados teve de ser normalizado seguindo os seguintes processos:

- ✓ foram excluídos os caracteres: < . > ; : ? / ° ~ ^ { [^ ' ` }] ° + = _) (* & " % \$ # @ ! " ' \ |
- ✓ foram excluídos os números;
- ✓ foram retirados acentos;
- ✓ foram excluídos os hífens;
- ✓ foram substituídas todas as letras maiúsculas por letras minúsculas.

A retirada dos acentos e dos hífens justifica-se porque, seja pelas perdas nos processos de conversão, seja por características originais dos textos, alguns fragmentos não possuíam hífens ou caracteres acentuados. O software *Frequência NumaBoa* contabiliza itens como “itaúna” e “itauna”, diferenciados somente pela presença do acento, como sendo diferentes. Dessa forma, deveria ser feita uma padronização de forma que o software contabilizasse tais itens como sendo o mesmo item. Essa normalização causou perdas na contabilização de itens diferenciados pelo acento na escrita, entretanto, essa perda não afetou de forma significativa os itens a serem analisados pela pesquisa.

O procedimento de normalização do arquivo de texto final foi feito com o auxílio do software *TexPad*. O arquivo final, em formato txt, normalizado segundo as regras descritas, continha um total de 1.997.942 palavras.

4.6.2.2. Contagem de frequência

Para a contagem de frequências de ocorrência dos itens utilizou-se o software *Frequência NumaBoa*. Esse software faz a contagem dos itens identificando-os como uma seqüência de letras entre espaços. O arquivo final gerado é uma lista de todos os itens ocorridos, seguidos da quantidade de vezes que cada um ocorreu no texto analisado em ordem crescente do total de suas ocorrências. Foram listadas 66.942 itens.

A localização dos itens de interesse da pesquisa pôde ser feita de forma automatizada a partir de um recurso disponível no software TextPad. As frequências de ocorrência de cada item e a ordenação decrescente das mesmas foram feitas pelo Microsoft Excel. A lista completa com as frequências dos itens terminados em /l/+vogal no corpus escrito encontra-se no Anexo 5.

4.7. Seleção das Variáveis independentes

4.7.1 Variáveis sociais

4.7.1.1. Gênero e Faixa Etária

Desde os primeiros estudos variacionistas, o gênero e a faixa etária têm sido considerados como fatores importantes na seleção de formas componentes de variáveis lingüísticas. Fisher (1974) [1958], em um estudo sobre o *ing* formador de gerúndio do inglês, observou que a variante velar, considerada a forma prestigiada, era mais freqüentemente utilizada na fala das crianças do sexo feminino. A partir desse estudo, Fisher concluiu que as formas mais prestigiosas deveriam ser mais escolhidas pelas mulheres. Desde então, diversos estudos variacionistas têm discutido a hipótese levantada por Fisher e, freqüentemente, encontrado uma relação direta entre gênero e prestígio social, identificando uma tendência das mulheres ao favorecimento de variantes de maior prestígio.

Sobre esse aspecto, Labov (1972) afirma que

In careful speech, women use fewer stigmatized forms than men, (Labov 1966a:288), and are more sensitive than men to the prestige pattern. They show this in a sharper slope of style shifting, especially at the more formal end of the spectrum. This observation is confirmed innumerable times, in Fisher (1958), throughout Shuy and Fasold's work in Detroit, in Levine and Crockett, and in Anshen's study of Hillsboro. The pattern is particularly marked in lower-middle-class women, who show the most extreme form of this behavior. (LABOV, 1972, p.243).

Por outro lado, Labov (1994) afirma que nos fenômenos de mudança “from below” (abaixo do nível da consciência social) a tendência maior é de que as mulheres estejam à frente do processo: “In linguistic change from below, women use higher frequencies of innovative forms than men do.”(LABOV, 1994, p. 292).

Interessa-nos, a partir deste estudo, identificar essa relação entre gênero e variação e mudança lingüística a partir da caracterização do fenômeno observado na fala dos moradores da cidade de Itaúna.

Em Martins (2001), em um estudo sobre o cancelamento das consoantes líquidas intervocálicas, a autora conclui que os fatores gênero e faixa etária parecem não exercer influência sobre o cancelamento das líquidas intervocálicas no dialeto falado na cidade de Belo Horizonte. Observações preliminares do dialeto falado na cidade de Itaúna puderam indicar um possível favorecimento da variável lingüística em questão relacionado a tais fatores. Observei-se que os homens pareciam tender à produção de formas mais reduzidas.

Observei-se também que a forma velarizada parece estar mais presente na fala dos homens jovens do que na fala dos demais estratos, talvez como uma marcação desse grupo social. Dessa forma, a faixa etária também parece incidir sobre as escolhas das variantes. A princípio, tinham-se indícios de que haveria um favorecimento de formas mais reduzidas pelos falantes da faixa etária jovem. Caso esse favorecimento fosse realmente observado, poderíamos avaliar se os fenômenos variáveis tratam-se também de mudança lingüística em progresso.

Entretanto, identificar padrões no uso das variantes diferenciados pela faixa etária pode ser apenas um indício de mudança em progresso. Para isso, seria necessária uma pesquisa da variação em tempo real, o que está além dos objetivos desse estudo. Ainda assim, identificar sincronicamente a variação favorecida pela idade pode oferecer bons indícios de um processo de mudança em progresso.

Outras variáveis sociais, como escolaridade, condição sócio-econômica, etnicidade ou região geográfica, não foram estratificadas nessa pesquisa, o que não tira delas a possibilidade de interferência na variável dependente. Aqui, busquei controlá-las ao máximo, para que os possíveis efeitos gerados por tais variáveis pudessem ser minimizados. Esse controle foi feito a partir da escolha de falantes homogeneamente distribuídos em relação às variáveis não consideradas.

4.7.1.2. Estilo

O estilo tem sido considerado, nos estudos sociolingüísticos, como uma importante variável descritiva. Considera-se essa variável tendo em vista o grau de formalidade da situação. Estilos mais formais fazem com que os falantes fiquem mais atentos à sua própria fala e faz com que eles alterem as formas variantes em função daquilo que eles consideram ser mais adequado para situações mais formais.

Dessa forma, a identificação de um aumento da sensibilidade do falante associado a estilos mais formais pode dar indícios do grau de consciência do fenômeno e se o mesmo é socialmente estigmatizado; assim, interessa-me observar a sensibilidade da variável ao estilo. Para tal, considere quatro estilos distintos: a) entrevista; b) perguntas direcionadas; c) leitura de texto e d) leitura de palavras.

4.7.2. Variáveis Internas

4.7.2.1. Contexto seguinte

Essa variável refere-se à característica fonética inicial do item seguinte ao item que apresenta /l/+vogal na sílaba final átona. Ao considerar essa variável, interessava-me observar a interferência do contexto fonético seguinte à realização de cada uma das formas variantes. A hipótese é de que a variação nos segmentos analisados em Itaúna pudesse ser favorecida foneticamente. Para tal, considere os contextos seguintes:

- 1 – consoante
- 2 – vogal
- 3 – pausa

Como pode ser observado no Anexo 13, uma divisão mais detalhada dos contextos seguintes não apresentou diferenças significativas em relação aos efeitos dos fatores. Assim, optei por uma recodificação mais simples dos fatores na variável contexto seguinte, o que gerou efeitos importantes associados a tal variável nos resultados finais.

4.7.2.2. Classe da palavra seguinte

Essa variável foi inserida na análise para que se pudesse avaliar a hipótese proposta em Bybee (2001) de que estruturas compostas por pronomes seguidos de verbos auxiliares, por serem estruturas muito freqüentes, seriam mais propensas à redução do que outras estruturas.

Yet it is undeniable that auxiliaries contract and fuse with subject pronouns, not with the following verb, even though it is the auxiliary and verb that belong to the same constituent. The reason is that specific instances of [pronoun + auxiliary] are extremely frequent, much more so than any particular sequences of [auxiliary + verb]. (BYBEE, 2001, p.186).

Considerarei também a divisão da variável em outros fatores para que isso pudesse ser observado mais detalhadamente. A variável classe morfológica da palavra seguinte ficou assim composta:

1. Verbo auxiliar
2. Verbo não-auxiliar
3. Não verbo
4. Não se aplica (pausa)

A inserção dessa variável também permite discutir as propostas desenvolvidas em Correa (1998) de que a redução nos pronomes ele(s) e ela(s), em posição de sujeito, está relacionada a um processo de cliticização de tais itens. Interessa-me observar se o fenômeno restringe-se a ambientes pré-verbais, como descrito em Correa (op.cit.), ou se o fenômeno é característico dos itens analisados por ele, independentemente de motivações sintáticas, em qualquer posição. Apesar de não haver correspondência exata entre o ambiente pré-verbal e a posição de sujeito, na maior parte dos dados, os pronomes ele (s) e ela (s), em posição pré-verbal, apresentam-se como sujeito.

4.7.2.3. Contexto anterior

A inclusão desta variável foi motivada para averiguar se a realização das variantes é favorecida pelo ambiente que precede a sílaba formada por l+vogal. Martins (2001) concluiu que a qualidade da vogal anterior parece não influenciar no cancelamento da lateral, mas apresenta indícios do apagamento ser maior, ainda que pouco significativo,

quando a vogal anterior é baixa. A variável contexto anterior é composta por:

1. Vogal alta ([i] e [u])
2. Vogal média ([e], [ɛ], [o] e [ɔ])
3. Vogal baixa ([ə])

4.7.2.4. Classe da palavra

Essa variável parece-nos fundamental na compreensão do fenômeno em análise. Em Correa (1998), o autor apresenta indícios que explicariam a variação na classe pronominal atribuindo tal variação a um processo de cliticização do pronome (ele/ela) na língua portuguesa. Diante disso, interessa-nos testar a hipótese de Correa tendo em vista classes de palavras diferentes das observadas por ele. Considerar a variável classe morfológica permite que uma avaliação mais apurada da abrangência do fenômeno variável seja realizada. Considerei os seguintes fatores para a variável *classe da palavra*:

1. Nome
2. Pronome
3. Verbo

4.7.2.5. Fricativa alveolar seguinte

A inclusão dessa variável foi feita com o objetivo principal de diferenciar formas de plural de formas de singular e avaliar a interferência de uma fricativa alveolar, representada pelo arquifonema /S/, adjunta à sílaba final analisada. Em Martins (2001), não se fazem diferenciações entre itens com /S/ ou itens sem /S/, entretanto, nas tabelas apresentadas pela autora, percebe-se uma diferença significativa, em termos percentuais, na comparação entre tais itens, conforme exposto a seguir:

Tabela 3: Cancelamento da lateral intervocálica com e sem presença de /S/ final em Martins (2001)

PALAVRAS	CANCELA		NÃO CANCELA		TOTAL
dele	5	56%	4	44%	9
deles	6	86%	1	14%	7
ela	34	20%	135	80%	169
elas	4	57%	10	43%	14
ele	26	12%	188	88%	214
eles	42	29%	102	71%	144
favela	3	20%	12	80%	15
favelas	2	67%	1	33%	3

Recorte da tabela apresentada em Martins (2001, p. 65)

A partir da tabela 3, pode-se ter indícios de que a presença da fricativa alveolar final parece favorecer de alguma forma o apagamento da lateral intervocálica.

4.7.2.6. Vogal da variável

Interessava-nos, com a inclusão dessa variável na análise, observar se a variação poderia ter alguma motivação associada à altura da vogal na sílaba final. A não-realização da vogal seguinte ao /l/, por exemplo, pode estar associada a um processo de redução vocálica favorecida pela altura das vogais i e u.

4.7.2.7. Tonicidade

O objetivo da inclusão desta variável neste estudo é identificar se os processos em análise são influenciados pela tonicidade. Interessa-nos analisar se os fenômenos de apagamento da vogal, de apagamento da sílaba e de velarização da lateral são favorecidos por características acentuais do item lexical em análise.

4.7.3. Outras Variáveis

Além das variáveis independentes apresentadas nas seções anteriores, analisarei também outras variáveis consideradas em estudos que discutem a variação e a mudança lingüística, mais recentemente. A inserção dessas variáveis na dicotomia *variáveis internas* e *variáveis sociais* é discutível. Por esse motivo, tais variáveis foram colocadas em um tópico separado.

4.7.3.1. Freqüência

A freqüência de ocorrência do item lexical e a freqüência de ocorrência de padrões sonoros desempenha, segundo Bybee (2001), um papel fundamental na mudança sonora, cada um a seu modo, conforme já discutido no capítulo 3. Assim, a inserção da variável freqüência na análise pode oferecer indicações de interferência da freqüência do item no fenômeno em análise. A possível associação entre freqüência e variação pode oferecer indícios da interferência da freqüência nos fenômenos de mudança sonora. A hipótese é de que itens mais freqüentes são atingidos primeiro em processos de etiologia fonética.

4.7.3.2. Indivíduo

Oliveira (1997), apoiando-se em Shen (1990), afirma que o comportamento do indivíduo não é uma reprodução do grupo ao qual ele pertence. Os resultados mostrados em Oliveira (op.cit.) indicam que “indivíduos de um mesmo grupo social podem apresentar comportamentos completamente contrários em relação ao cancelamento do (r) final em nominais” (p. 56). A inserção dessa variável na análise busca avaliar a proposta de Oliveira (1997) e identificar padrões individuais que possam diferenciar-se de uma estratificação social nos moldes da sociolinguística variacionista (gênero, faixa etária, classe social, escolaridade, procedência e grupo étnico). Podemos ainda aventar a hipótese de termos outros agrupamentos atuando, como aquelas propostos nos estudos das redes sociais (*social network.s*), segundo Milroy, L (1987).

4.7.3.3. Item lexical

A variável item lexical foi inserida para que pudessem ser discutidas e avaliadas as propostas difusionistas que consideram que a mudança se dá item por item, ou grupos de itens por grupos de itens, e que, diferentemente do que propõem as correntes neogramáticas, uma mudança sonora apresenta seleção lexical. Para a análise variacionista, Oliveira (1992) sugere que “o comportamento individual seja checado para todos os itens lexicais” (p. 40). Uma análise dessa natureza implica a atribuição de códigos específicos para cada indivíduo e para cada item lexical. Além disso, a atribuição de códigos diferenciados para cada um dos itens possibilita retirá-los ou inseri-los facilmente na

análise, caso haja a suspeita de que a motivação da variação seja muito influenciada por alguns itens específicos.

A inserção da variável *item lexical* também permite que sejam avaliados os efeitos de frequência na variação a partir da distribuição das variantes em relação a cada item individualmente. A codificação foi feita a partir da palavra, os verbos *falo* e *fala* foram codificados separadamente.

CAPÍTULO 5

MODELOS DE ANÁLISE QUANTITATIVA

A análise quantitativa está na base dos estudos sociolinguísticos variacionistas. Desde os primeiros estudos em variação linguística (Fisher (1974) [1958], Labov (1963) e Labov (1964)) têm-se adotado modelos quantitativos para dar suporte à necessidade de se estudar a linguagem em uso e a variação linguística na forma como ela ocorre na fala de uma comunidade linguística. A identificação da variabilidade ocorrida nas línguas como um fenômeno dependente de determinadas variáveis, e não como uma “variação livre” como propunham os estruturalistas, foi possível a partir do desenvolvimento de técnicas eficazes de análise de tal variabilidade. A homogeneidade do sistema linguístico e a ocorrência da “variação livre” na perspectiva estruturalista deram lugar, nos estudos variacionistas, à heterogeneidade, passível de observação e de quantificação, e a uma variabilidade favorecida por fatores sociais e por fatores linguísticos.

A partir de uma pesquisa bibliográfica na área da sociolinguística variacionista, constatei que o aprofundamento no estudo dos métodos estatísticos utilizados no estudo da variação linguística não tem recebido, na maioria das vezes, a atenção merecida. Acredito que tal aprofundamento seja de fundamental importância, na medida em que grande parte dos estudos variacionistas utiliza metodologias de análise quantitativa.

O software Varbrul, criado por Cedergren e Sankoff no início da década de 70, e aprimorado nos anos seguintes, permitiu aos linguistas interessados no estudo da variação o acesso a métodos de análise estatística. Entretanto, a facilidade no acesso e na interpretação dos resultados gerados pelo Varbrul gerou também a possibilidade de utilização de métodos estatísticos sem que se tivesse obrigatoriamente um conhecimento estatístico mais geral ou mesmo um conhecimento mais específico dos procedimentos internos utilizados pelo próprio Varbrul. Guy (1998, p. 25), num texto intitulado *Varbrul: análise avançada*, afirma que os métodos e problemas associados à análise quantitativa, entre linguistas, "(...) costumam passar de boca em boca, uma versão acadêmica da tradição pré-letrada da história oral". O autor afirma que escreve "não como um desbravador ou inovador, mas como um mero compilador das histórias contadas por aqueles que sabem". As idéias expressas por Guy revelam uma realidade nos estudos variacionistas. Uma busca por

referências bibliográficas mais aprofundadas do modelo estatístico utilizado pelo Varbrul e, conseqüentemente, do modelo mais utilizado pela SV, indica que todo o desenvolvimento teórico do modelo estatístico implantado no Varbrul parece concentrar-se em uma pequena quantidade de textos. Além disso, o Varbrul parece carregar consigo o mito de que somente ele é capaz de analisar com eficiência os dados coletados nos estudos variacionistas e que o método estatístico e os procedimentos realizados por ele são exclusivos e específicos para a análise lingüística. Nessa dissertação tentarei mostrar que isso não é verdade, ou pelo menos, não é totalmente verdade. O Varbrul utiliza um modelo estatístico que é amplamente difundido e utilizado nas mais diversas áreas do conhecimento, entretanto, ele diferencia-se por utilizar métodos pouco convencionais de codificação⁶ das variáveis independentes.

Acredito que limitarmos o estudo da estatística ao estudo dos procedimentos do Varbrul e de suas referências bibliográficas leva-nos a uma limitação no desenvolvimento metodológico do estudo da variação lingüística, na medida em que a compreensão mais aprofundada de outros modelos e do próprio modelo de regressão logística pode possibilitar análises mais amplas dos fenômenos em estudo. Por esse motivo, nesse capítulo apresentarei um aprofundamento do modelo estatístico utilizado nos estudos variacionistas, bem como uma interpretação mais clara das especificidades do Varbrul em relação aos métodos convencionais de estimação de parâmetros. Utilizarei um estilo de escrita mais didático, pois acredito que esse texto possa ser útil para outros interessados no aprofundamento dos métodos estatísticos utilizados na sociolingüística variacionista. A não-familiaridade com uma linguagem estatística talvez dificulte um pouco a leitura em alguns trechos, por isso, tentarei explicar os procedimentos da forma mais clara possível. Aprofundamentos com relação à teoria estatística podem ser obtidos em Pagano e Gauvreau (2004), Hosmer e Lemeshow (2000), Kleinbaum (1994), entre outros.

Algumas questões que motivaram este estudo serão respondidas no decorrer do capítulo. Entre elas:

1. O que diferencia o Varbrul de outros softwares estatísticos convencionais?
2. O que é e como é calculado matematicamente o peso relativo?
3. O Varbrul é o único software capaz de manipular com eficiência a especificidade dos dados variacionistas?

⁶ Este termo refere-se a um procedimento automático realizado pelo software, independente dos códigos atribuídos pelo pesquisador aos fatores nas variáveis independentes e às variantes da variável dependente.

Para a análise do banco de dados utilizarei os softwares GoldVarbX (de agora em diante *Varbru*) e o SPSS v.13.0 (de agora em diante *SPSS*). O banco de dados analisado nos exemplos foi criado a partir de dados reais de fala, coletados na cidade de Itaúna/MG. Utilizarei na maior parte dos exemplos a variável dependente formada pelas variantes [l]⁺vogal e Ø ocorridas na sílaba final átona de itens lexicais terminados em /l/⁺vogal, como em *ele ~ ê* e *aquela ~ aquê*.

5.1 O modelo estatístico

A seleção de um modelo⁷ estatístico a ser utilizado se dá, primeiramente, a partir de quais perguntas se quer responder. Nos estudos em SV, relacionados à variação sonora, tem-se que um som ora é produzido de uma forma, ora é produzido de outra forma, em um mesmo item. Nesse caso, a pergunta é: o que poderia estar influenciando os indivíduos de uma mesma comunidade a falarem uma mesma palavra ora de uma forma, ora de outra? Os modelos estatísticos que permitem responder a esta pergunta, ou seja, que permitem que se possa explicar a variabilidade de um fenômeno em relação um conjunto de fatores, são chamados de modelos de regressão. Nos modelos de regressão temos sempre uma variável, chamada variável dependente ou variável resposta, e uma ou mais variáveis explicativas, chamadas de covariáveis, variáveis independentes ou grupo de fatores, que poderão ajudar a explicar a variabilidade na variável resposta.

A seleção do modelo estatístico a ser utilizado também depende do tipo de variável resposta do estudo. Se a variável resposta fosse contínua, por exemplo, poderíamos optar um modelo de regressão linear. Estudos utilizando tal modelo podem ser encontrados, por exemplo, em Labov (1994) e Labov (2001). Esse modelo poderia ser utilizado caso a variável resposta fosse, por exemplo, a frequência dos formantes de uma vogal, no qual teríamos observações localizadas em uma faixa contínua. Por outro lado, se a variável resposta fosse categórica, poderíamos utilizar o modelo de regressão logística. Esse modelo poderia ser utilizado caso a variável resposta fosse composta de somente duas possibilidades, por exemplo, a presença ou a ausência da marcação de plural em itens nominais. Se a variável resposta fosse categórica e apresentasse mais de duas possibilidades,

⁷ Nesta seção utilizarei os termos *modelo* e *método* da forma como eles são normalmente utilizados na literatura estatística. Neste contexto, *modelo* será utilizado para fazer referência à equação da regressão logística com as variáveis independentes selecionadas; o termo *método* será utilizado para fazer referência, na maioria das vezes, à forma como as variáveis são codificadas.

poderíamos utilizar o modelo logístico multinomial. Tal modelo poderia ser utilizado, por exemplo, no estudo do pronome você, caso fossem consideradas as realizações de mais de duas variantes, considerados por hipótese como categóricas, como você, ocê e cê.

Na maior parte dos estudos em SV tem-se utilizado o modelo de regressão logística, já que é esse o modelo implementado no *Varbrul*. Esse modelo é utilizado quando se quer investigar, dado um conjunto de possíveis variáveis independentes, quais delas estão de fato associadas à variável dependente.

Tomemos o conjunto de dados de fala no qual encontramos duas possibilidades para os itens lexicais terminados em sílaba átona formada por /l/+vogal, dadas por a) [l]+vogal e b) \emptyset . Assim, itens como *aquele* e *tranquilo*, poderiam ocorrer também como *aquê* e *tranqüii*. Nesse caso, temos uma variável dependente binária [l]+vogal e \emptyset .

Suponha que tenhamos a hipótese de que o gênero influencie a produção de uma ou outra forma. Assim, temos uma variável independente *gênero*, composta pelos fatores⁸ *masculino* e *feminino*.

A variável dependente apresenta duas possibilidades, ou ocorre [l]+vogal ou ocorre \emptyset . Podemos codificá-la então como 0 ou 1. Nesse caso, determinei que a variante codificada como 0 seja a variante [l]+vogal e a variante codificada como 1 seja a variante \emptyset , já que meu objetivo é analisar o fenômeno de apagamento da sílaba final /l/+vogal. Em geral, a variante codificada como 1 será aquela que se quer investigar. Normalmente, essa variante de referência recebe o nome de *sucesso*, em estatística, e de “variante não-padrão”, “aplicação da regra variável”, entre outros, nos estudos variacionistas.

Se utilizarmos o modelo de regressão logística para a análise da variável [l]+vogal (0) e \emptyset (1) tendo como variável independente o gênero (masculino ou feminino) poderemos determinar a influência do gênero na probabilidade de sucesso da variável dependente, ou seja, na probabilidade de utilização da variante \emptyset . Além disso, o modelo permite que seja analisado simultaneamente o efeito de múltiplas variáveis independentes. Nas seções seguintes, o modelo de regressão logística será aprofundado.

⁸ Utilizarei o termo *fator* para fazer referência ao que se denomina categoria na literatura estatística.

5.2. Definições preliminares

5.2.1. Hipótese nula, nível de significância e p-valor

Qualquer hipótese levantada para explicar estatisticamente algum fenômeno vem associada a uma segunda hipótese que nega a primeira. Tais hipóteses são chamadas, respectivamente, de hipótese alternativa e hipótese nula.

Nos modelos de regressão, por exemplo, um teste estatístico poderia propor testar a hipótese nula de não haver efeito na variável dependente associado a uma variável independente. Assim, a suposição de que a variabilidade em itens lexicais terminados em /l/+vogal possa ser explicada pelo gênero dos falantes vem acompanhada da hipótese nula que sugere que o gênero dos falantes não exerce influência estatisticamente significativa sobre a variabilidade nos itens. A hipótese efetivamente testada é a hipótese nula.

A probabilidade máxima aceitável de rejeitarmos a hipótese nula quando ela é de fato verdadeira é chamada de *nível de significância*. No exemplo acima, o nível de significância seria a probabilidade máxima de aceitarmos que o gênero do falante interfere na variabilidade, quando na realidade ele não interfere. O *nível de significância* é um valor arbitrário, definido segundo critérios do pesquisador. Convencionalmente, na sociolinguística variacionista, assim como em outras ciências, utilizamos um *nível de significância* de 0,05.

Todo teste estatístico clássico fornece a probabilidade do efeito observado ser proveniente do acaso. Tal probabilidade é chamada de *p-valor*. Um modo de conduzir um teste estatístico, é o de rejeitar a hipótese nula quando o *p-valor* é menor que o nível de significância. Nesse caso, se encontramos um *p-valor* de 0,02 em um teste estatístico, podemos afirmar que a hipótese nula foi rejeitada, já que o *p-valor* foi menor do que o nível de significância de 0,05. Isso indica que os resultados obtidos são estatisticamente significantes.

5.2.2. Probabilidade, chance e razão de chances

A probabilidade pode ser definida como uma medida numérica da possibilidade de ocorrência de um evento qualquer em uma população. Na prática, ela pode ser obtida pela razão entre o número de ocorrências de um evento e o número total de ocorrências da amostra. Vejamos a tabela de contingência⁹ abaixo, obtida a partir da classificação de uma amostra de 2280 observações segundo o gênero e a variável dependente [1]+vogal ~ Ø:

Tabela 4: Tabela de Contingência para o gênero na variável [1]+vogal ~ Ø no estilo entrevista

	Masculino	Feminino	Total
[1]+vogal	356	715	1071
Ø	588	621	1209
Total	944	1336	2280

A partir dos dados acima podemos calcular a probabilidade de um indivíduo utilizar a variante Ø, bem como a probabilidade associada aos gêneros:

$$\hat{p}_{total} = \frac{1209}{2280} = 0,53$$

$$\hat{p}_{masc} = \frac{588}{944} = 0,62$$

$$\hat{p}_{fem} = \frac{621}{1336} = 0,46$$

sendo \hat{p}_{total} a probabilidade total de ocorrência de Ø, \hat{p}_{masc} a probabilidade de Ø dado que o indivíduo pertence ao gênero masculino e \hat{p}_{fem} a probabilidade de Ø dado que o indivíduo pertence ao gênero feminino.

Como a probabilidade é sempre um número entre 0 e 1, se temos a probabilidade de ocorrer um evento (p), a probabilidade de que tal evento não ocorra será 1-p. Assim, a probabilidade de ocorrer Ø em um indivíduo do gênero masculino é 0,62 e a probabilidade de não ocorrer Ø é 1-0,62=0,38.

⁹ Tabela de contingência é uma tabela de cruzamento de dados de duas variáveis categóricas.

Outra medida importante é a *chance*, ou *odds*, definida como a razão entre a probabilidade de que um evento ocorra e a probabilidade de que ele não ocorra. Assim, a *chance* para as probabilidades acima é dada por:

$$odds_{total} = \frac{0,53}{1-0,53} = 1,13$$

$$odds_{masc} = \frac{0,62}{1-0,62} = 1,65$$

$$odds_{fem} = \frac{0,46}{1-0,46} = 0,87$$

A interpretação de um resultado em termos de chance é feita da seguinte forma: a probabilidade de ocorrer \emptyset entre os homens é 1,65 vezes a probabilidade de não ocorrer \emptyset , ou seja, de ocorrer [1]+vogal; e a probabilidade de ocorrer \emptyset entre as mulheres é 0,87 vezes a probabilidade de ocorrer [1]+vogal. A partir desses resultados podemos inferir que a chance de ocorrer a variante \emptyset entre os homens é quase duas vezes a chance de ocorrer entre as mulheres (1,65 para 0,87). A comparação entre as chances de ocorrência de um evento entre fatores de uma variável é chamada de *razão de chances* e é bastante utilizada nos modelos de regressão logística. A razão de chances, ou *odds ratio*, fornece uma medida do grau de associação entre fatores de uma variável independente. No exemplo acima, a razão de chances entre o gênero masculino e o gênero feminino é obtida através da razão entre as *chances* dos gêneros,

$$OR = \frac{odds_{masc}}{odds_{fem}} = \frac{1,65}{0,87} = 1,9$$

A razão de chances de 1,9 indica que a chance de *sucesso* (nesse caso, a variante \emptyset) para o gênero masculino é 1,9 vezes a chance no gênero feminino. Isso indica que a chance de um homem, na cidade de Itaúna, utilizar a variante \emptyset é quase duas vezes a chance de uma mulher utilizar a mesma variante.

5.3. Modelo de regressão logística

O modelo de regressão logística é utilizado quando temos uma variável dependente binária, nos casos dos exemplo acima, [1]_{+vogal} ou \emptyset , e variáveis independentes que podem contribuir para explicarmos a variabilidade na variável dependente. Matematicamente, o modelo de regressão logística, considerando n variáveis independentes, é definido pela equação:

$$\log \frac{p}{(1-p)} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Na equação acima,

- ✓ a função $\log \frac{p}{(1-p)}$, ou simplesmente $\text{logit}(p)$, é chamada de função de ligação;
- ✓ p é a probabilidade de *sucesso*;
- ✓ α é a constante que expressa o valor do $\text{logit}(p)$ quando todas as variáveis independentes são iguais a 0;
- ✓ β (variando de 1 a n) é um coeficiente que expressa o efeito das variáveis independentes x (variando de 1 a n) na função de ligação, quando a variável x aumenta uma unidade.

Os valores de α e dos β 's são estimados pelo método da máxima verossimilhança e obtidos a partir de um algoritmo numérico computacional¹⁰. Esse método encontra, entre todos os valores possíveis, os valores de α e dos β 's que sejam mais prováveis de terem gerado os dados observados.

Tomando, por exemplo, como variável dependente [1]_{+vogal} (codificada como 0) e \emptyset (codificada como 1) e como variável independente o gênero - feminino (0) e masculino (1) - a equação da regressão logística estimada será definida por:

¹⁰ Maiores informações em Dobson (1990) e McCullagh & Nelder (1989).

$$\log \frac{\hat{p}}{(1-\hat{p})} = \hat{\alpha} + \hat{\beta}_1 \text{gênero}_1$$

onde \hat{p} é a estimativa da probabilidade de ocorrência de variante \emptyset , $\hat{\alpha}$ é a constante e $\hat{\beta}_1$ o efeito da variável *gênero* no $\text{logit}(\hat{p})$, quando ela aumenta uma unidade, ou seja, passa de 0 (feminino) para 1 (masculino).

Rodando no SPSS a regressão logística, tendo como variável dependente [1]+vogal (0) e \emptyset (1) e como variável independente o gênero, teremos o seguinte valor para β (expresso na saída do SPSS para o modelo acima):

$$\hat{\beta}_1 = 0,643$$

Portanto, temos um aumento de 0,643 no $\text{logit}(\hat{p})$, quando passamos do gênero feminino (0) para o gênero masculino (1).

Utilizando-se propriedades do logaritmo e tomando-se a exponencial do coeficiente β , obtemos uma interpretação da associação em termos de razão de chances:

$$OR = \exp(\beta)$$

Assim, dado que $\hat{\beta}$ para a variável gênero é 0,643, a razão de chances entre os fatores da variável gênero será:

$$\hat{OR} = \exp(0,643) = 1,9$$

O valor 1,9 para a razão de chances corresponde ao mesmo valor encontrado na seção anterior. Entretanto, o modelo exemplificado aqui contém somente uma variável independente. Em um modelo multivariado, a estimativa do efeito de uma variável altera com a inserção de outras variáveis no modelo, o que não permite que a razão de chances seja obtida das chances calculadas a partir de uma tabela de contingência, como na seção 5.2.2. Nesse caso, temos uma razão de chances que leva em consideração o efeito das demais variáveis independentes.

5.3.1. Seleção das variáveis independentes

Após definirmos o conjunto de variáveis independentes a serem incluídas no modelo de regressão logística, precisamos identificar as variáveis mais importantes para explicar a probabilidade de sucesso. Hosmer e Lemeshow (2000) afirmam que, tradicionalmente, a construção de um modelo estatístico implica na busca do modelo mais parcimonioso para a explicação dos dados¹¹. Assim, interessa-nos, entre todas as variáveis independentes consideradas, identificar o conjunto de variáveis que melhor contribui para explicarmos a variabilidade.

O processo de seleção de variáveis pode ser feito de maneiras diversas. Os métodos utilizados no *Varbrul*, chamados *step-up* e *step-down*, são conhecidos em estatística como métodos *stepwise*, respectivamente como *forward* e *backward*, e são encontrados nos softwares estatísticos convencionais. No SPSS, eles recebem o nome de *forward-lr* e *backward-lr*.

Os métodos *stepwise* permitem que a seleção das variáveis independentes mais importantes seja feita de forma automática, segundo critérios pré-definidos. Sobre os métodos *stepwise*, Hosmer e Lemeshow (2000) afirmam:

Any stepwise procedure for selection or deletion of variables from a model is based on a statistical algorithm that checks for the "importance" of variables, and either includes or excludes them on the basis of a fixed decision rule. The "importance" of a variable is defined in terms of a measure of the statistical significance of the coefficient for the variable. (HOSMER E LEMESHOW, 2000, p. 16).

O critério de seleção utilizado nesses métodos, em ambos os softwares, é a *razão da máxima verossimilhança*. No método *step-up* ou *forward-lr*, os programas inserem, passo a passo, cada uma das variáveis independentes, uma a uma, e comparam o valor do *logaritmo da verossimilhança*¹² ou *log-likelihoods* dos modelos com e sem a variável independente adicionada.

No passo 1, obtém-se o valor do *log-likelihood* sem a inclusão de nenhum parâmetro associado às variáveis independentes, ou seja, considera-se somente o parâmetro $\hat{\alpha}$. No passo seguinte, testam-se todas as variáveis inseridas no modelo uma a uma e seleciona-se somente aquela que gera o maior valor no *teste da razão da máxima verossimilhança*.

¹¹ The traditional approach to statistical model building involves seeking the most parsimonious model that still explain the data. (Hosmer e Lemeshow, 2000, p. 92)

¹² Uma *função de verossimilhança* é uma função que fornece a probabilidade de obter os dados observados, dados os valores dos parâmetros. O *logaritmo da verossimilhança* é uma transformação de uma *função de verossimilhança* usando logaritmos naturais.

O procedimento se repete nos passos seguintes. As variáveis selecionadas em cada passo vão sendo mantidas no modelo utilizado para seleção da próxima variável. A seleção das variáveis independentes é interrompida quando a diferença entre o modelo sem a variável e o modelo com a variável não apresenta significância estatística (p -valor $< 0,05$)¹³.

Os métodos *step-up* e *backward-LR* selecionam as variáveis mais importantes tendo como referência um modelo em que todas as variáveis são incluídas. Em cada passo seguinte, testa-se cada uma das variáveis e retira-se aquela que apresenta o menor valor no *teste da razão da máxima verossimilhança*, ou seja, que apresenta o maior *p*-valor. A retirada de variáveis é interrompida quando o teste apresenta resultado estatisticamente significativo

A seleção de variáveis pelos métodos *stepwise* é feita de forma automática pelos softwares estatísticos. A utilização desses métodos gera um modelo que inclui somente as variáveis mais importantes para a explicação do fenômeno em análise. Esse procedimento facilita bastante o trabalho do pesquisador, entretanto, a seleção baseia-se exclusivamente em critérios estatísticos.

No SPSS, o método padrão de inserção de variáveis no modelo é o método manual, denominado *enter*. No *Varbrul*, esse procedimento é chamado de *one-level*. O método *enter* ou *one-level* permite que a entrada das variáveis independentes seja feita de forma manual, segundo critérios do pesquisador. A partir da comparação das saídas de modelos contendo conjuntos diferentes de variáveis independentes, utilizando o *teste da razão da máxima verossimilhança*, pode-se determinar o melhor conjunto de variáveis para explicar a probabilidade de sucesso. Esse método é bastante interessante, mas exige do pesquisador um maior domínio dos métodos de comparação de modelos para que o melhor modelo possa ser eficientemente selecionado.

¹³ O *p*-valor para o *teste da razão da máxima verossimilhança* pode ser obtido a partir de uma tabela de distribuição do *qui-quadrado* com *n* graus de liberdade, sendo *n* o número de variáveis incluídas no segundo modelo.

5.4. Comparando o *Varbrul* e o *SPSS*

Como veremos nessa seção, os efeitos estimados das variáveis independentes apresentados pelo *Varbrul* diferem-se dos efeitos apresentados em uma saída padrão do *SPSS*. Como será mostrado, o resultado gerado pelo *Varbrul* diferencia-se pela maneira como são codificados os fatores que compõem as variáveis independentes. Essa diferença gera efeitos diferenciados para as variáveis. No *SPSS*, assim como em outros softwares estatísticos, podemos definir o parâmetro de codificação dos fatores de maneira semelhante à codificação feita internamente pelo *Varbrul* e, assim, obtermos resultados semelhantes.

5.4.1. Diferentes formas de codificação de fatores

5.4.1.1. Fator de referência

Numa regressão logística convencional utiliza-se uma codificação de fatores das variáveis independentes em que determina-se um fator de referência. Tal codificação é amplamente difundida e utilizada nas mais diversas áreas do conhecimento. No *SPSS*, assim como em outros pacotes estatísticos, a codificação com fator de referência é dada como padrão. Nela, os efeitos dos demais fatores de uma variável independente e, conseqüentemente, a razão de chances, estarão em relação ao fator de referência. No exemplo mostrado na seção 5.3, o fator de referência¹⁴ para a variável gênero foi o gênero *feminino*. Assim, a razão de chances obtida (1,9) refere-se à chance do gênero *masculino* (1,9) em relação ao gênero *feminino* (1).

A codificação dos fatores é feita pelo software a partir de uma tabela, chamada *matriz de desenho*. Quando temos somente dois fatores, a *matriz de desenho*, tendo o primeiro fator como referência, é:

Tabela 5: Matriz de desenho para a variável gênero

Gênero	Código
Feminino	0
Masculino	1

¹⁴ A escolha do fator de referência na variável independente é feita pelo pesquisador e não afeta os resultados. Se o fator de referência fosse o gênero masculino, a razão de chances seria 0,53, obtido dividindo-se 1 (feminino) por 1,9 (masculino).

No caso de termos uma variável com três fatores, a matriz de desenho, tendo o primeiro fator como referência, é:

Tabela 6: Matriz de desenho para a variável contexto seguinte

Contexto Seguinte	Código do parâmetro (1)	Código do parâmetro (2)
Consoante	0	0
Vogal	1	0
Pausa	0	1

Nesse caso, a equação do modelo de regressão logística é dado por

$$\log \frac{\hat{p}}{(1-\hat{p})} = \hat{\alpha} + \hat{\beta}_1 \text{vogal} + \hat{\beta}_2 \text{pausa}$$

Os efeitos estimados, obtidos a partir de um modelo de regressão logística tendo como variável dependente $[1]_{+vogal}$ e \emptyset e como variável independente os contextos seguintes *consoante*, *vogal* e *pausa*, são:

$$\hat{\alpha} = 0,546$$

$$\hat{\beta}_1 = -1,034$$

$$\hat{\beta}_2 = -2,187$$

A razão de chances entre os fatores, como mostrado na seção 5.3 é o exponencial do $\hat{\beta}$. Assim, teremos as seguintes razões de chances para a variável *contexto seguinte*, tendo o fator *consoante* como fator de referência:

$$OR_{vogal} = \exp(-1,034) = 0,36$$

$$OR_{pausa} = \exp(-2,187) = 0,11$$

A chance de um indivíduo utilizar a variante \emptyset antes de vogal é 0,36 a chance de utilizá-la antes de consoante. Da mesma forma, a chance de um indivíduo utilizar a variante \emptyset antes de pausa é 0,11 a chance de utilizá-la antes de consoante. As relações entre as OR das categorias não definidas como referência também podem ser feitas. Assim, a chance de um indivíduo utilizar a variante \emptyset antes de vogal é 3,27 (0,36/0,11) a chance de utilizá-la antes de pausa.

5.4.1.2. Desvio da média

O modelo de regressão logística implantado no *Varbrul* é idêntico ao modelo implantado em outros pacotes estatísticos. Entretanto, ele difere-se pela maneira como os fatores das variáveis independentes são codificados. Em vez de tomar um fator como referência e, a partir dele, estimar o efeito dos demais fatores, o *Varbrul* utiliza uma codificação conhecida como *desvio da média*. No método *desvio da média*, a variável *gênero* é codificada como (-1) *feminino* e (1) *masculino*, ao invés de 0 e 1, respectivamente, como no método *fator de referência*. A diferença na codificação gera diferentes efeitos estimados, como pode ser visto a seguir.

Tomando o conjunto de dados com variável dependente [1]+vogal e \emptyset e variável independente o contexto seguinte *consoante*, *vogal* e *pausa*, temos a seguinte tabela de contingência:

Tabela 7: Tabela de contingência para o contexto seguinte na variável dependente apagamento da sílaba /l/+vogal

	Consoante	Vogal	Pausa	Total
[1]+vogal	603	184	284	1071
\emptyset	1041	113	55	1209
Total	1644	297	339	2280

A partir da tabela 7, podemos obter a chance de cada fator. O método *desvio da média* utiliza o logaritmo (ln) da chance dos fatores (\hat{g}), dado por:

$$\hat{g}_{cons} = \ln\left(\frac{1041}{603}\right) = 0,546$$

$$\hat{g}_{vog} = \ln\left(\frac{113}{184}\right) = -0,488$$

$$\hat{g}_{pansa} = \ln\left(\frac{55}{284}\right) = -1,642$$

A partir dos valores acima, podemos calcular também um valor médio ($\hat{g}_{\text{média}}$) dos fatores:

$$\hat{g}_{\text{média}} = \frac{\hat{g}_{\text{cons}} + \hat{g}_{\text{vog}} + \hat{g}_{\text{pausa}}}{3} = -0,528$$

Os valores dos efeitos estimados ($\hat{\beta}$) a serem incluídos na equação da regressão logística, no método *desvio da média*, são dados pelos \hat{g} 's das categorias subtraídos da média dos \hat{g} 's, ou seja,

$$\hat{\beta}_1 = \hat{g}_{\text{cons}} - \hat{g}_{\text{média}} = (0,546) - (-0,528) = 1,074$$

$$\hat{\beta}_2 = \hat{g}_{\text{vog}} - \hat{g}_{\text{média}} = (-0,488) - (-0,528) = 0,040$$

$$\hat{\beta}_3 = \hat{g}_{\text{vogal}} - \hat{g}_{\text{média}} = (-1,642) - (-0,528) = -1,114$$

A partir dos *betas* calculados acima, podemos calcular uma razão de chances em relação à média (OR') para cada categoria:

$$\text{OR}'_{\text{cons}} = \exp(1,074) = 2,926$$

$$\text{OR}'_{\text{vog}} = \exp(0,040) = 1,041$$

$$\text{OR}'_{\text{pausa}} = \exp(-1,114) = 0,328$$

A razão de chances proveniente do método *desvio da média* não é a mesma da razão de chances calculada pelo método *fator de referência*. Hosmer e Lemeshow (2000, p. 60) afirmam que tal *razão de chances* é uma razão de chances do fator em relação à *média geométrica* das chances de todos os fatores da variável independente¹⁵.

Diante dos resultados apresentados acima, como obter o *factor weights* ou *peso relativo*, fornecido na saída do *Varbrul*? De acordo com Morrison (2005), o *peso relativo* é dado por

¹⁵ Exponentiation of the estimated coefficients yields the ratio of the odds for the particular group to the geometric mean of the odds. (Hosmer e Lemeshow, 2000, p. 60)

$OR'/(1+OR')$, ou seja, o *peso relativo* é uma medida calculada a partir da razão de chances, calculada pelo método *desvio da média*. Como a razão de chances é um número entre 0 e infinito, o *peso relativo* será sempre um número entre 0 e 1. Caso a OR' seja igual a 1, teremos que o *peso relativo* será igual a 0,50. Assim, os *pesos relativos* para as categorias da variável *contexto seguinte* são:

$$PR_{cons} = \frac{2,926}{1 + 2,926} = 0,745$$

$$PR_{vog} = \frac{1,041}{1 + 1,041} = 0,510$$

$$PR_{pausa} = \frac{0,328}{1 + 0,328} = 0,247$$

O valor do *input* corresponde ao *peso relativo* da média das *chances*, dado por:

$$input = \frac{\exp(\hat{g}_{media})}{1 + \exp(\hat{g}_{media})} = \frac{\exp(-0,528)}{1 + \exp(-0,528)} = 0,371$$

Os resultados acima podem ser comprovados pelos resultados obtidos por uma rodada com *fatores centralizados* no *Varbrul*:

```
Run # 2, 3 cells:
Convergence at Iteration 5
Input 0.371
Group # 1 -- C: 0.745, V: 0.510, P: 0.247
Log likelihood = -1428.074 Significance = 0.000
```

De acordo com Rand e Sankoff (1990), a opção *fatores centralizados*¹⁶ no *Varbrul* considera que os fatores em um grupo possuem pesos iguais; se não selecionarmos essa opção, cada fator recebe um peso de acordo com a sua ocorrência relativa no total de ocorrências no grupo.

Se multiplicamos os valores das frequências relativas encontrados na tabela abaixo, pelos valores encontrados para os \hat{g} 's dos fatores, temos um valor médio dos \hat{g} 's

¹⁶ When this option is chosen, each factor in a group is given equal weight. Otherwise each factor is weighted according to its occurrences relative to total occurrences of all factors in the group. Hand e Sankoff (1990)

considerando a frequência relativa das ocorrências em cada fator particular. Assim, temos:

Tabela 8: Tabela de contingência para variável dependente apagamento da sílaba /l/+vogal com frequência relativa dos fatores da variável contexto seguinte

	Consoante	Vogal	Pausa	Total
[l]+vogal	603	184	284	1071
Ø	1041	113	55	1209
Total	1644	297	339	2280
Freq. relativa	0,72	0,13	0,15	1,0

$$\hat{g}_{\text{cons}} \times 0,72 = 0,546 \times 0,72 = 0,393$$

$$\hat{g}_{\text{vog}} \times 0,13 = -0,488 \times 0,13 = -0,063$$

$$\hat{g}_{\text{pausa}} \times 0,15 = -1,642 \times 0,15 = -0,246$$

O valor para $\hat{g}_{\text{média}}$ é dado pela soma das multiplicações dos \hat{g} 's dos fatores pelas frequências de suas ocorrências em relação ao número total de ocorrências do grupo, ou seja, a média agora é ponderada em relação às frequências relativas de cada fator:

$$\hat{g}_{\text{média}} = 0,393 + (-0,063) + (-0,246) = 0,084$$

A partir dos valores acima podemos recalculer os valores dos efeitos estimados ($\hat{\beta}$) para cada fator:

$$\hat{\beta}_{\text{cons}} = 0,546 - 0,084 = 0,462$$

$$\hat{\beta}_{\text{vog}} = -0,488 - 0,084 = -0,572$$

$$\hat{\beta}_{\text{pausa}} = -1,642 - 0,084 = -1,726$$

Podemos recalcular também suas razões de chances, dadas por:

$$OR'_{\text{cons}} = \exp(0,462) = 1,587$$

$$OR'_{\text{vog}} = \exp(-0,572) = 0,564$$

$$OR'_{\text{pausa}} = \exp(-1,726) = 0,178$$

A partir das OR' de cada fator, aplicando a fórmula $PR = OR' / (1 + OR')$, os pesos relativos são:

$$PR_{\text{cons}} = 1,587 / (1 + 1,587) = 0,613$$

$$PR_{\text{vog}} = 0,564 / (1 + 0,564) = 0,360$$

$$PR_{\text{pausa}} = 0,178 / (1 + 0,178) = 0,151$$

Rodando novamente os dados no *Varbrul*, desmarcando a opção *fatores centralizados*, obteremos exatamente os mesmos valores obtidos acima:

```
Run # 2, 3 cells:
Convergence at Iteration 5
Input 0.522
Group # 1 -- C: 0.613, V: 0.360, P: 0.151,
Log likelihood = -1428.074 Significance = 0.000
```

Os resultados apresentados acima indicam que, de fato, o *Varbrul* utiliza um método diferenciado de parametrização dos efeitos dos parâmetros. Tal método assemelha-se ao *desvio da média*, mas o efeito do parâmetro é estimado considerando-se a magnitude da interferência do fator na variável dependente a partir da quantidade de ocorrências do mesmo. A justificativa para a utilização de um método específico, apresentada em Sankoff (1988), é de que os dados coletados nos estudos em variação linguística diferem-se por apresentar uma distribuição desigual nos dados, como ocorre, por exemplo, na tabela 8, em que temos 72% das ocorrências da variável no fator *consoante* e 13% e 15% nos fatores *vogal* e *pausa*, respectivamente.

5.4.2. Comparando as saídas

A constatação de que o método utilizado pelo *Varbrul* parece ser específico dele, leva-nos a outro questionamento, relevante do ponto de vista metodológico: Por que outras áreas do conhecimento, especialmente nas ciências sociais, que também contam com dados mal distribuídos, não utilizam um método que considere o *desvio da média* com uma média ponderada a partir dos efeitos dos fatores? A resposta a esta questão exige um aprofundamento maior em termos de conhecimentos estatísticos para que se avalie as implicações na utilização de tal método e estão além dos objetivos deste trabalho. Entretanto, podemos testar aqui as implicações práticas em termos de resultados.

Tomemos como exemplo um modelo que tenha como variável dependente as variantes [l]⁺vogal e Ø e como variáveis independentes: *gênero, idade, contexto seguinte, altura da vogal, tipo de informação no turno, contexto anterior, classe da palavra seguinte, tonicidade, presença de /S/ no final da palavra e classe da palavra.*

O resultado abaixo foi gerado a partir de uma saída padrão de regressão logística do SPSS, com variáveis significativas selecionadas pelo método *forward-lr*:

Tabela 9: Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [L]+vogal ~ Ø, no estilo entrevista, utilizando o SPSS¹⁷

Variáveis Independentes	Fatores	n ₁ / n _r	% ₁	p-valor (Wald)	Razão de chances	
Gênero	Feminino	621 / 1336	46,5	<0,001	0,39	
	Masculino*	588 / 944	62,3		1,00	
Faixa Etária	Adulto	532 / 1012	52,6	0,008	0,76	
	Jovem*	677 / 1268	53,4		1,00	
Contexto Seguinte	Pausa	55 / 339	16,2	<0,001	0,12	
	Consoante*	1041 / 1644	63,3		1,00	
	Vogal	113 / 297	38,0		<0,001	0,34
Classe da Palavra	Nome*	46 / 335	13,7	<0,001	1,00	
	Pronome	1147 / 1855	61,8		6,34	
	Verbo	16 / 90	17,8		0,048	2,01
Classe da Palavra Seguinte	Verbo auxiliar*	110 / 143	76,9	<0,001	1,00	
	Verbo não-aux.	657 / 965	68,1		0,43	
	Não-verbo	372 / 814	45,7		<0,001	0,33
	Pausa	70 / 358	19,6		0,031	0,39
Vogal na variável	Vogal A	329 / 909	36,2	0,678	0,87	
	Vogal U*	17 / 115	14,8		1,00	
	Vogal I	863 / 1256	68,7		0,019	2,19
Tonicidade	Proparoxítona	13 / 23	56,5	<0,001	12,1	
	Paroxítona*	1196 / 2257	53,0		1,00	
/S/ final da palavra	Ausente*	877 / 1816	48,3	<0,001	1,00	
	Presente	332 / 464	71,6		1,89	

* fator de referência

Na tabela acima, os valores expresso na coluna *razão de chances* representam a razão de chances entre os fatores das variáveis independentes. Na variável gênero, por exemplo, a chance de uma mulher utilizar a variante Ø é 0,39 vezes a chance de um homem utilizar tal variante. Para a variável contexto seguinte, a chance de ocorrer Ø antes de pausa é 0,12 vezes a chance de ocorrer Ø antes de consoante; a chance de ocorrer Ø antes de vogal é 0,34 vezes a chance de ocorrer Ø antes de consoante. Pode-se obter a razão de chances entre os fatores dividindo-se uma razão de chances pela outra. O fator que ficar no numerador será o fator de referência. Assim, dividindo-se 0,34 por 0,12 temos que a chance de ocorrer Ø antes de vogal é 2,8 vezes a chance de ocorrer Ø antes de pausa. A coluna *p-valor (Wald)* refere-se ao p-valor obtido a partir do teste de Wald. A partir deste teste é

¹⁷ Saída completa do SPSS no Anexo 8.

possível determinar a significância dos fatores dentro de uma variável independente, comparativamente ao fator tomado como fator de referência. Um p-valor $>0,05$ no teste de *Wald*, para um determinado fator, permite-nos dizer que tal fator não apresenta efeito estatisticamente diferente do fator de referência. Para casos em que a variável independente apresenta somente dois fatores, teremos a significância entre eles. Entretanto, quando temos mais de dois fatores em uma variável independente, a análise da significância no teste de *Wald* precisa ser realizada com cautela, na medida em que a escolha do fator de referência não é motivada, ou seja, qualquer fator pode ser tomado como referência (as razões de chances mantêm sempre as mesmas proporções). Dizer, por exemplo, que, na variável *vogal na variável*, o efeito do fator *vogal A* não é estatisticamente diferente do fator *vogal U*, não é o mesmo que dizer que o fator *vogal A* não apresenta significância estatística, já que o efeito de tal fator pode ser estatisticamente diferente do fator *vogal I*. Assim, a significância dada pelo teste de *Wald* em variáveis com mais de dois fatores, diferentemente da Razão de Chances, muda dependendo do fator de referência escolhido, já que a significância é em relação ao fator de referência. Na variável independente *vogal na variável*, o fator *vogal A* apresentou p-valor=0,678, isto significa que o efeito do fator *vogal A* não apresenta diferença estatisticamente significativa em relação ao efeito do fator de referência *vogal U*, mas não significa que o efeito do fator *vogal A* não seja significativo. Se o fator de referência escolhido fosse, por exemplo, a *vogal I*, teremos as seguintes significâncias: I: p-valor $<0,001$; A: p-valor $<0,001$ e U: p-valor=0,029 (as razões de chances mantêm as mesmas proporções). Esses valores indicam que os efeitos dos fatores *vogal A* e *vogal U* são estatisticamente diferentes do fator de referência *vogal I*. Diante disso, na análise dos resultados, a significância do fator em variáveis independentes com mais de dois fatores será analisada em relação a todos os demais fatores que compõem a variável.

O resultado abaixo foi obtido de uma saída do *Varbrul*, com variáveis significativas selecionadas pelo método step-up. A rodada 44 é aquela que contém o melhor modelo, ou seja, aquela que contém as variáveis mais importantes.

Tabela 10: Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ Ø, no estilo entrevista, utilizando o Varbrul¹⁸

Variáveis Independentes	Fatores	n ₁ / n _t	% ₀₁	Peso Relativo
Gênero	Feminino	621 / 1336	46,5	0,40
	Masculino	588 / 944	62,3	0,64
Faixa Etária	Adulto	532 / 1012	52,6	0,46
	Jovem	677 / 1268	53,4	0,53
Contexto Seguinte	Pausa	55 / 339	16,2	0,16
	Consoante	1041 / 1644	63,3	0,61
	Vogal	113 / 297	38,0	0,35
Classe da Palavra	Nome	46 / 335	13,7	0,18
	Pronome	1147 / 1855	61,8	0,58
	Verbo	16 / 90	17,8	0,30
Classe da Palavra Seguinte	Verbo auxiliar	110 / 143	76,9	0,71
	Verbo não-aux.	657 / 965	68,1	0,52
	Não-verbo	372 / 814	45,7	0,45
	Pausa	70 / 358	19,6	0,48
Vogal na Variável	Vogal A	329 / 909	36,2	0,37
	Vogal U	17 / 115	14,8	0,41
	Vogal I	863 / 1256	68,7	0,60
Tonicidade	Proparoxítona	13 / 23	56,5	0,92
	Paroxítona	1196 / 2257	53,0	0,49
/S/ final da Palavra	Ausente	877 / 1816	48,3	0,47
	Presente	332 / 464	71,6	0,62

log-likelihood: -1176.058; p-valor = 0,009

A tabela abaixo apresenta uma comparação dos resultados utilizando-se o SPSS (*categoria de referência*), resultados expressos em *razão de chances*, e o Varbrul (*desvio da média “ponderada”*), resultados expressos em *peso relativo*.

¹⁸ Saída completa do Varbrul no Anexo 16

Tabela 11: Comparação entre as saídas do Varbrul e do SPSS para a variável dependente [L]+vogal ~ Ø

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Peso Relativo	Razão de chances	
Gênero	Feminino	621 / 1336	46,5	<0,001	0,40	0,39	
	Masculino*	588 / 944	62,3		0,64	1,00	
Faixa Etária	Adulto	532 / 1012	52,6	0,008	0,46	0,76	
	Jovem*	677 / 1268	53,4		0,53	1,00	
Contexto Seguinte	Pausa	55 / 339	16,2	<0,001	0,16	0,12	
	Consoante*	1041 / 1644	63,3		0,61	1,00	
	Vogal	113 / 297	38,0	<0,001	0,35	0,34	
Classe da Palavra	Nome*	46 / 335	13,7	<0,001	0,18	1,00	
	Pronome	1147 / 1855	61,8		0,58	6,34	
	Verbo	16 / 90	17,8		0,048	0,30	2,01
Classe da Palavra Seguinte	Verbo auxiliar*	110 / 143	76,9	<0,001	0,71	1,00	
	Verbo não-aux.	657 / 965	68,1		0,52	0,43	
	Não-verbo	372 / 814	45,7		<0,001	0,45	0,33
	Pausa	70 / 358	19,6		0,031	0,48	0,39
Vogal na variável	Vogal A	329 / 909	36,2	0,678	0,37	0,87	
	Vogal U*	17 / 115	14,8		0,41	1,00	
	Vogal I	863 / 1256	68,7		0,019	0,60	2,19
Tonicidade	Proparoxítona	13 / 23	56,5	<0,001	0,92	12,1	
	Paroxítona*	1196 / 2257	53,0		0,49	1,00	
/S/ final da palavra	Ausente*	877 / 1816	48,3	<0,001	0,47	1,00	
	Presente	332 / 464	71,6		0,62	1,89	

* fator de referência

Os resultados acima indicam que a seleção das variáveis estatisticamente significativas em ambos os softwares foi a mesma, ou seja, as variáveis independentes selecionadas e excluídas do modelo foram idênticas. Vemos também que a direção da influência do fator também é a mesma, já que uma ordenação dos fatores das variáveis a partir do efeito gerado na variável dependente é também a mesma. Além disso, as diferenças em termos de peso relativo entre os fatores de uma variável apresentam semelhanças em relação às diferenças em termos de razão de chances. Se tomarmos por exemplo as variáveis gênero e faixa etária, vemos que a diferença entre os efeitos dos fatores nas variáveis em termos de peso relativo (masculino=0,64 para feminino=0,40 e jovem=0,53 para adulto=0,46) corresponde à diferença dos efeitos em termos de razão de chances (masculino=1,00 para feminino=0,39 e jovem=1,00 para adulto=0,76). Isso parece indicar que a utilização de um método em que se mede o *desvio da média "ponderada"* a partir

dos efeitos dos fatores, como o utilizado pelo Varbrul, não apresenta diferença significativa em relação a um método convencional com *fator de referência*, como no SPSS, ou em outros pacotes estatísticos, ainda que os dados sejam mal distribuídos, como no caso dos dados utilizados para gerar os resultados acima.

Para efeito de comprovação dos resultados apresentados, apresentarei abaixo o mesmo procedimento, incluindo as mesmas variáveis independentes ao modelo, mas com as demais variáveis dependentes.

Variável dependente: [1]+vogal ~ [1]

Apagamento da vogal

Tabela 12: Comparação entre as saídas do Varbrul¹⁹ e do SPSS²⁰ para a variável dependente [1]+vogal ~ [1]

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Peso Relativo	Razão de chances
Gênero	Feminino	430 / 1145	37,6	<0,001	0,38	0,33
	Masculino*	563 / 919	61,3		0,65	1,00
Contexto Seguinte	Pausa	191 / 475	40,2	<0,001	0,25	0,26
	Consoante*	486 / 1089	44,6		0,53	1,00
	Vogal	316 / 500	63,2		<0,001	0,69
Classe da Palavra Seguinte	Verbo auxiliar*	41 / 74	55,4	0,351	0,50	1,00
	Verbo não-aux.	325 / 633	51,3		0,43	0,78
	Não-verbo	413 / 855	48,3		0,406	0,80
	Pausa	214 / 502	42,6		0,022	0,68
Vogal na variável	Vogal A	257 / 837	30,7	<0,001	0,30	0,26
	Vogal U*	161 / 259	62,2		0,63	1,00
	Vogal I	575 / 968	59,4		0,717	0,64
/S/ final da palavra	Ausente*	942 / 1881	50,1	<0,001	0,53	1,00
	Presente	51 / 183	27,9		0,25	0,29

* fator de referência

¹⁹ Saída completa do Varbrul no Anexo 14

²⁰ Saída completa do SPSS no Anexo 6

**Variável dependente: [1] ~ [±]
Velarização da lateral**

Tabela 13: Comparação entre as saídas do Varbrul²¹ e do SPSS²² para a variável dependente [1] ~ [±]

Variáveis Independentes	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Peso Relativo	Razão de chances	
Gênero	Feminino	21 / 451	4,7	<0,001	0,32	0,29	
	Masculino*	100 / 663	15,1		0,62	1,00	
Faixa Etária	Adulto	35 / 512	6,8	<0,001	0,42	0,55	
	Jovem*	602 / 86	14,3		0,57	1,00	
Contexto Seguinte	Pausa	73 / 264	27,7	0,031	0,75	4,53	
	Consoante*	24 / 510	4,7		0,39	1,00	
	Vogal	24 / 340	7,1		0,394	0,45	1,30
Classe da Palavra	Nome*	26 / 247	10,5	<0,001	0,45	1,00	
	Pronome	93 / 817	11,4		0,58	2,93	
	Verbo	2 / 50	4,0		0,509	0,39	0,60
Classe da Palavra Seguinte	Verbo auxiliar*	1 / 42	2,4	0,590	0,24	1,00	
	Verbo não-aux.	14 / 339	4,1		0,35	1,77	
	Não-verbo	34 / 447	7,6		0,141	0,59	4,60
	Pausa	72 / 286	25,2		0,202	0,59	4,86
Tonicidade	Proparoxítona	3 / 6	50,0	0,005	0,95	18,4	
	Paroxítona*	118 / 1108	10,6		0,50	1,00	

* fator de referência

Os resultados apresentados acima indicam que o SPSS e o Varbrul, apesar de utilizarem métodos diferenciados de codificação de fatores, geram resultados semelhantes do ponto de vista analítico, ainda que os dados da amostra não sejam bem distribuídos. A diferença principal é que o SPSS, além do *teste da razão da máxima verossimilhança*, fornece a significância no *teste de Wald*, a partir do qual é possível identificar se o efeito de um fator é estatisticamente diferente do efeito de outro fator, em uma mesma variável independente.

²¹ Saída completa do Varbrul no Anexo 15

²² Saída completa do SPSS no Anexo 7

5.4.3. Interação entre variáveis independentes

Em algumas situações, uma variável independente pode interagir com outra variável independente; isso ocorre quando o efeito de um fator em uma variável dependente está relacionado ao efeito de outro fator em outra variável independente. A variável *gênero*, por exemplo, pode ter um efeito diferente na variante \emptyset da variável dependente [l]_{+vogal} ~ \emptyset , dependendo do valor da variável *faixa etária*. Isso pode acontecer, por exemplo, se os falantes *adultos* do gênero *masculino* tiverem um comportamento diferente dos falantes *jovens* do gênero *masculino* com relação ao uso da variante \emptyset . Se isso ocorrer, o efeito do fator *masculino* na variável gênero será alterado, caso o fator da variável faixa etária seja alterado. Com relação a fatores lingüísticos, isso poderia ocorrer, por exemplo, se o efeito da *vogal seguinte* ao /l/ na variante \emptyset fosse alterado quando a *classe da palavra* fosse alterada. Suponha, por exemplo, que o efeito da vogal seguinte [l], fosse diferente em pronomes e em verbos.

Essa interferência no efeito de uma variável gerada por outra variável é chamada de *interação*. Diversos autores têm discutido os problemas gerados pela interação nos dados sociolingüísticos, como Sankoff (1988), Guy (1988), Saito (1999), Bayley (2002), Morrison (2005), entre outros. Em comum, esses autores discutem a mesma limitação do Varbrul: a impossibilidade ou a dificuldade de ajustes do modelo com dados que apresentam interação.

Em Guy (1988), o autor afirma que

A análise do Varbrul difere de alguns outros tipos de análises multivariadas pelo fato de pressupor que os vários grupos de fatores têm efeitos independentes. Mas os usuários deveriam compreender claramente que eles não devem pressupor essa independência. (GUY, 1988, p.41)

Nesse artigo, Guy expõe dois procedimentos propostos em Cerdegren (1973) para que sejam determinadas as variantes que poderiam estar interagindo. O primeiro procedimento seria observar, na saída do Varbrul, valores altos para o p-valor localizado no final de cada célula. Caso existam vários valores altos envolvendo co-ocorrência repetida de um par de variáveis, provavelmente a causa seria a interação. O segundo procedimento seria agrupar duas variáveis que pudessem estar interagindo (criando uma única variável unindo gênero e idade, por exemplo) e comparar as análises de duas rodadas: uma com as variáveis agrupadas e outra com as variáveis não agrupadas. Se houver uma ordem nos pesos dos fatores na variável agrupada diferente da ordem esperada construída através das

duas variáveis separadas, então há possibilidade de interação. Esse procedimento pode ser visto também em Sankoff (1988).

Saito (1999), Bayley (2002) e Morrison (2005) afirmam que o Varbrul não consegue determinar a interação entre duas variáveis em um conjunto de dados e que tal procedimento precisaria ser realizado a partir de outro software estatístico. Em Guy (1988) e Sankoff (1988), os autores afirmam que os procedimentos adotados no Varbrul podem supor que haja interação nos dados, mas não podem prever se determinada interação existe ou não entre duas variáveis independentes e, principalmente, determinar se a interação é significativamente relevante.

No SPSS, podemos identificar facilmente se há variáveis interagindo no conjunto de variáveis independentes. Tal procedimento é feito a partir de um recurso que multiplica uma variável x_1 por uma variável x_2 para que se forme uma nova variável x_1x_2 . As variáveis x_1 e x_2 estarão interagindo caso o p-valor da multiplicação de tais variáveis, no teste de *Wald*, seja significativo, ao se inserir a variável criada no modelo de regressão. A variável criada a partir da multiplicação de outras variáveis é chamada de termos da interação.

Tomando como exemplo os dados apresentados na tabela 12, vemos que a variável faixa etária foi excluída do modelo de regressão ao utilizarmos um método de seleção stepwise para a variável dependente $[1]_{+vocal} \sim [1]$.

Se o *termo da interação* entre gênero e faixa etária for incluído ao modelo, vemos que a faixa etária passa a ser significativa e que o termo da interação também é estatisticamente significativo. A tabela abaixo é um recorte da saída do SPSS apresentada no Anexo 6 com todas as variáveis independentes selecionadas por stepwise e com inclusão de um termo de interação entre gênero e faixa etária.

Tabela 14: Recorte do modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [1]+vogal ~ [1], no estilo entrevista, incluindo o termo da interação entre gênero e faixa etária²³

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	430 / 1145	37,6	<0,001	0,2
	Masculino	563 / 919	61,3		1,0
Faixa Etária	Adulto	477 / 957	49,8	0,009	0,6
	Jovem	516 / 1107	46,6		1,0
Termo de Interação	Gênero _F x Faixa Etária _A			<0,001	2,7

Os resultados apresentados na tabela 14 indicam que o termo da interação é estatisticamente significativo (p-valor < 0,05) e que, portanto, gênero e faixa etária interagem, no conjunto de dados analisados. Se o termo da interação não fosse considerado, poderíamos concluir que a variável faixa etária não apresentou resultados estatisticamente significativos nos dados analisados o que, de fato, parece não ser verdade. A inclusão da variável *faixa etária* em um modelo que considera a interação entre gênero e faixa etária indica-nos que tais fatores sociais devem ser analisados com mais cautela, na medida em que um fator parece exercer influência em outro fator. A interação e a análise das variáveis gênero e faixa etária serão apresentadas nas seções subseqüentes.

5.4.4. Avaliação dos softwares

As discussões apresentadas neste capítulo permitem que cheguemos a algumas conclusões em relação aos softwares utilizados. Nessa seção, serão apresentados tópicos de avaliação do Varbrul e do SPSS. Tentarei também apontar resumidamente os pontos positivos e os pontos negativos de cada um deles. As questões que inicialmente motivaram um estudo estatístico mais aprofundado, expostas no início do capítulo, puderam ser respondidas no decorrer do capítulo e serão sumarizadas nos tópicos abaixo.

- ✓ Como mostrado nesse capítulo, o modelo estatístico implantado no Varbrul é um modelo amplamente utilizado e disponível em outros pacotes estatísticos, chamado de modelo de regressão logística. Entretanto, o método de codificação dos fatores

²³ Saída completa do SPSS no Anexo 6.

das variáveis independentes é diferenciado no Varbrul. Normalmente, o método de codificação padrão dos demais pacotes estatísticos é o método *fator de referência*, o *Varbrul* utiliza um método chamado *desvio da média*. O método *desvio da média* também é encontrado na maioria dos pacotes estatísticos; no SPSS, tal método é chamado de contraste *deviation*. Entretanto, o *Varbrul* utiliza um tipo especial de *desvio da média*. Ao invés de obter uma média simples a partir da soma dos efeitos dos fatores dividida pelo número de fatores, o *Varbrul* obtém uma média “ponderada”, obtida a partir da soma da multiplicação de cada fator pela sua frequência relativa em relação a todos os fatores.

- ✓ Na seção 5.4.3 vimos que o *Varbrul* não lida muito bem com a interação entre variáveis independentes. Vimos também que considerar o termo da interação, desde que ele seja significativo, altera os efeitos das variáveis e, portanto, a interação deve ser considerada. No entanto, a análise da interação entre os fatores requer conhecimentos estatísticos mais aprofundados.
- ✓ O fato do *Varbrul* limitar-se ao modelo de regressão logística com variável dependente binária e variáveis independentes categóricas traz limitações ao pesquisador. A utilização de um pacote estatístico mais completo possibilita que sejam avaliadas situações em que a variável dependente possui mais de 2 categorias (uso de um modelo logístico multinomial) ou em que a variável independente seja contínua (uso de um modelo de regressão linear).
- ✓ Por utilizar um método muito específico de codificação e estimação dos efeitos, o *Varbrul* limita bastante o leque de opções de materiais disponíveis para compreender seus procedimentos internos. Além disso, minha visão pessoal é de que a linguagem utilizada nos textos que explicam os procedimentos do *Varbrul* é bastante obscura, o que dificulta um paralelo com outros textos estatísticos. Ao contrário, a utilização de um método convencional de codificação das variáveis é amplamente discutido nos textos estatísticos e pode ser encontrado em qualquer material que trate do modelo logístico.
- ✓ A especificidade do *Varbrul* com relação à codificação e estimação dos efeitos das variáveis não gera resultados significativamente diferentes em relação ao método padrão utilizado pelo SPSS, como mostrado na seção 5.4.2. Isso indica que utilizar o *Varbrul* ou qualquer outro pacote estatístico que possua regressão logística não altera

os resultados de estudos em sociolinguística variacionista, mesmo se os dados forem mal distribuídos. A utilização do SPSS, entretanto, fornece mais informações, como por exemplo, a significância entre os fatores de uma variável independente.

- ✓ Como pontos positivos, o SPSS apresenta ainda a possibilidade de criação de gráficos e tabelas, a fácil manipulação do banco de dados e a compatibilidade com outros softwares. Como pontos negativos ressalto o fato do SPSS ser um software proprietário de alto custo e de exigir uma capacidade alta de processamento do computador.
- ✓ Como pontos positivos, temos que o Varbrul é um software amplamente utilizado nos estudos em variação linguística, apresenta resultados familiares aos pesquisadores da área, é bastante simples e não exige muita capacidade de processamento do computador, além de ser um software de uso gratuito.
- ✓ Diversos softwares estatísticos poderiam ser utilizados na análise variacionista, basta que o software rode modelos de regressão logística. Entre os softwares disponíveis, destaco o R²⁴, que é um software gratuito e de código aberto. No site do programa afirma-se que o “R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible”. Há, inclusive, um projeto chamado R-Varb²⁵, desenvolvido por John Paolillo, no qual as rotinas do Varbrul são implementadas no R.

Tendo em vista os tópicos apresentados acima, optei pela utilização do SPSS. Ainda que não tenha sido possível avaliar, do ponto de vista teórico, os efeitos de se considerar o método *desvio da média ponderada* ou o método *fator de referência*, opto pelo método *fator de referência* por ser ele o método padrão do modelo de regressão logística e porque a escolha de um ou outro método não traz alterações significativas em termos de resultados.

²⁴ Informações e download do software em <http://www.r-project.org/>

²⁵ Informações disponíveis em <http://ella.slis.indiana.edu/~paolillo/>

CAPÍTULO 6

ANÁLISE DOS RESULTADOS

Neste capítulo serão analisados os resultados obtidos para as três variáveis dependentes conforme definidas na seção 2.2. Os resultados serão apresentados a partir da interferência de cada uma das variáveis independentes, apresentadas na seção 4.7, nas variáveis dependentes. Entre todos os estilos, a distribuição das variantes é:

Tabela 15: Distribuição das variantes nos estilos entrevista, perguntas direcionadas, leitura de texto e leitura de palavras

Variantes		Entrevistas	Perguntas Direcionadas	Leitura de Texto	Leitura de Palavras	Total
[l]+vogal	n	1071	108	1419	941	3539
	%	31,6	55,38	73,14	86,57	76,6
[l]	n	993	82	463	143	1681
	%	29,3	42,05	23,87	13,16	21,35
[ɫ]	n	121	4	18	2	145
	%	3,6	2,05	0,93	0,18	0,74
∅	n	1209	1	40	1	1251
	%	35,6	0,51	2,06	0,09	1,3
Total	n	3394	195	1940	1087	6616
	%	51,3	2,9	29,3	16,4	100,0

A variável *estilo*, quando inserida no modelo juntamente com as demais variáveis, apresentou significância nos três fenômenos analisados nesta pesquisa. Entretanto, para a análise que se segue, utilizarei os dados coletados no estilo *entrevista*, por ser ele o estilo que mais se aproxima da fala cotidiana. A análise da variável estilo será feita separadamente. Assim, serão analisados um total de 3.394 ocorrências, assim distribuídas:

Tabela 16: Distribuição das variantes no estilo *entrevista*

Variantes	n	%
[1]+vogal	1071	31,6
[1]	993	29,3
[ʎ]	121	3,6
Ø	1209	35,6
Total	3394	100,0

A quantidade de dados utilizados na análise de cada variável dependente considerada neste estudo será dada pela soma das ocorrências das variantes que compõe cada fenômeno variável. Segue a distribuição dos dados para os três fenômenos em análise:

Tabela 17: Distribuição dos dados para as variáveis dependentes no estilo *entrevista*

Variáveis dependentes		n
Apagamento da vogal	[1]+vogal ~ [1]	2064
Apagamento da vogal e velarização da lateral	[1] ~ [ʎ]	1114
Apagamento da Sílabas	[1]+vogal ~ Ø	2280

6.1. Teste de interações

Conforme discutido na seção 5.4.3, a interação entre variáveis independentes pode causar um falseamento dos efeitos dos fatores na variável dependente. Sankoff (1988) afirma que

Empirical studies of speech communities which are relatively homogeneous socially have tended not to find such interaction; with some exceptions, age and sex differentiation in phonology, morphology and syntax is generally quantitative rather than qualitative, and it is only by incorporating data from ethnically distinct groups or geographically separate communities in a study that we can expect to find ‘different grammars’. (SANKOFF, 1988, p. 992).

Com relação à interação entre fatores lingüísticos, o autor afirma que interações desse tipo freqüentemente são decorrentes de problemas na codificação dos fatores ou na formulação das variáveis. Por outro lado, “interaction among extralinguistic factors, however, is quite usual.” (Sankoff, 1988, p. 992).

Morrison (2005), analisando a variação na fricativa dental vozeada na fala do primeiro ministro canadense, considerou as variáveis *estilo de fala*, *segmento seguinte* e *vogal anterior*. Os testes de interação foram realizados a partir da inserção dos *termos da interação* no modelo, gerados a partir da multiplicação das variáveis, duas a duas. O autor testou também a interação entre as três variáveis simultaneamente. Segundo Morrison (op.cit.), nenhuma interação entre os grupos de fatores foi encontrada.

Saito (1999), reanalisando os dados de Beebe (1983) sobre os efeitos de variáveis independentes na aquisição de segunda língua de crianças porto-riquenhas, testou a interação de três variáveis consideradas em Beebe (op.cit.), da mesma maneira em que foi testada a interação em Morrison (2005): incluindo no modelo os termos da interação criados a partir da multiplicação das variáveis duas a duas e um termo da interação em que foram multiplicadas as três variáveis simultaneamente. O autor mostra que alguns efeitos apresentados em Beebe (1983) foram alterados pelo efeito de variáveis que interagem e que, portanto, deveriam ser reanalisados incluindo-se o termo de interação no modelo de regressão.

Guy (1988), analisando o apagamento do /s/ final no português do Brasil, identificou uma interação entre duas variáveis independentes de natureza lingüística: o ponto de articulação e o vozeamento da consoante seguinte. O autor constatou que o efeito do segmento velar seguinte era alterado pelo vozeamento de tal segmento. O segmento /g/ apresentou PR=.12 e o segmento /k/ apresentou PR=.27, ou seja, o segmento /k/ seguinte tinha um efeito maior do que o segmento /g/, sendo que tais segmentos diferenciavam-se apenas pela variável vozeamento. A conclusão para essa diferença, já que, segundo o autor, não foi encontrada nenhuma explicação razoável para ela, foi atribuir a diferença nos efeitos à má distribuição dos dados no corpus (/k/=961 e /g/=75) e à constatação de que a hipótese nula, de que o vozeamento não influenciava no apagamento do /s/ final, não deveria ter sido rejeitada.

Os exemplos acima mostram que a interação entre variáveis independentes pode gerar efeitos diferentes na variável dependente em análise e que, portanto, é importante que tais interações sejam identificadas.

Ao testarmos a interação das variáveis independentes lingüísticas em nosso banco de dados constatamos que a variável *classe da palavra* interage com a variável *contexto seguinte*, quando consideramos a variável dependente [l]_{+vogal} ~ ∅. Observando a distribuição das

variáveis independentes em relação à variável dependente numa tabela de contingência, temos:

Tabela 18: Tabela de contingência para a variável dependente
[l]+vogal ~ Ø com cruzamento das variáveis *contexto seguinte* e *classe da palavra*

Classe da Palavra			Contexto Seguinte			Total
			Consoante	Vogal	Pausa	
nome	[l]+vogal	n	148	40	101	289
		%	86,6	85,1	86,3	86,3
	Ø	n	23	7	16	46
		%	13,45	14,9	13,7	13,7
	Total	n	171	47	117	335
		% do total	51,0	14,0	34,9	100,0
pronome	[l]+vogal	n	418	130	160	708
		%	29,3	56,8	80,8	38,2
	Ø	n	1010	99	38	1147
		%	70,7	43,2	19,2	61,8
	Total	n	1428	229	198	1855
		% do total	77,0	12,4	10,7	100,0
verbo	[l]+vogal	n	37	14	23	74
		%	82,2	66,7	95,8	82,2
	Ø	n	8	7	1	16
		%	17,8	33,3	4,2	17,8
	Total	n	45	21	24	90
		% do total	50,0	23,3	26,7	100,0

A partir da tabela acima, temos que

- ✓ a probabilidade de ocorrer Ø em pronomes é 0,53 quando seguidos de consoante ($\frac{70,7\%}{70,7\% + 43,2\% + 19,2\%}$);
- ✓ a probabilidade de ocorrer Ø em verbos é 0,32 quando seguidos de consoante ($\frac{17,8\%}{17,8\% + 33,3\% + 4,2\%}$);
- ✓ a probabilidade de ocorrer Ø em nomes é 0,32 quando seguidos de consoante ($\frac{13,5\%}{13,5\% + 14,9\% + 13,7\%}$).

O resultado acima leva-nos à conclusão de que nos pronomes apaga-se mais a sílaba final quando seguidos de consoantes do que em outros itens quando também seguidos de consoantes. Não encontrei motivações fonético-fonológicas para esse resultado.

A opção por fazer testes de interação utilizando o mesmo método utilizado em Saito (1999) e Morrison (2005), em que todas as possibilidades de interação entre variáveis independentes fossem incluídas no modelo de regressão, teríamos que criar um modelo com um grande número de variáveis, já que em nosso estudo consideramos 9 variáveis independentes e, para criar os termos de interação, elas deveriam ser multiplicadas duas a duas, três a três, etc. O modelo final poderia ser reduzido a partir do método stepwise, em que somente as variáveis estatisticamente significantes seriam selecionadas. Isso poderia ser feito de forma relativamente simples no SPSS, conforme pode ser visto no Anexo 17. Entretanto, como conclui Guy (1988), nem sempre uma interação é gerada pelo efeito real entre variáveis independentes. Tal efeito pode ser atribuído à má distribuição dos dados analisados, como no exemplo acima, ou por algum outro fator identificável somente na reanálise das ocorrências. Isso significa que cada uma das interações deveria ser analisada individualmente, para que se chegasse à conclusão de que o efeito de uma variável é influenciado pelo efeito de outra variável com a qual ela interage ou por outro motivo qualquer; como a ocorrência de dados mal distribuídos.

A complexidade da análise da interação caso a caso, como mostrada acima, leva-nos à necessidade de fazermos escolhas de ordem prática. Por razões de tempo, é inviável que sejam analisadas todas as interações encontradas nos dados. Por outro lado, não considerar a interação, pode levar a interpretações que não refletem a realidade. Analisarei somente as interações ocorridas entre fatores sociais gênero e faixa etária, como sugere Sankoff (1988). Tal análise será apresentada nas seções seguintes.

6.2. Interação entre gênero e faixa etária

Nesta seção, mostrarei as possibilidades de interação entre fatores sociais gênero e faixa etária considerando os três fenômenos encontrados em Itaúna: apagamento da vogal, velarização e apagamento da sílaba. Os resultados e a discussão da influência de cada umas dessas variáveis será feita na seção seguinte. Aqui, discutirei somente a interferência da interação a partir de uma análise comparativa entre um modelo que considere variáveis interagindo e um modelo que não considere tal interação. Caso a interação seja confirmada, os termos da interação serão incluídos ao modelo final de regressão.

6.2.1. Apagamento da vogal – [1]+vogal ~ [1]

A tabela abaixo apresenta um recorte do resultado da regressão tendo como variável dependente [1]+vogal ~ [1], com inclusão de todas as variáveis independentes, selecionadas por stepwise, desconsiderando a possibilidade de interação:

Tabela 19: Saída do SPSS, com variável dependente [1]+vogal ~ [1] e variáveis independentes gênero e faixa etária²⁶

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	430 / 1145	37,6	<0,001	0,33
	Masculino*	563 / 919	61,3		1,00
Faixa Etária	Adulto	477 / 957	49,8	-	-
	Jovem*	516 / 1107	46,6	-	-

* fator de referência

Nessa tabela apresenta-se a razão de chances para a variável *gênero*. A razão de chances para a variável *faixa etária* não foi exibida pois ela não apresentou significância estatística, como pode ser visto no Anexo 6.

A tabela abaixo apresenta o resultado da regressão tendo como variável dependente [1]+vogal e [1], com inclusão de todas as variáveis independentes, selecionadas por stepwise, e inclusão do termo da interação entre gênero e faixa etária:

Tabela 20: Saída do SPSS, com variável dependente [1]+vogal ~ [1] e variáveis independentes gênero, faixa etária e o termo da interação entre gênero e faixa etária²⁷

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	430 / 1145	37,6	<0,001	0,21
	Masculino*	563 / 919	61,3		1,00
Faixa Etária	Adulto	477 / 957	49,8	0,001	0,62
	Jovem*	516 / 1107	46,6		1,00
Termo da Interação	Gênero _F x Faixa Etária _A			<0,001	2,6

* fator de referência

²⁶ Saída completa do SPSS no Anexo 6.

²⁷ Saída completa do SPSS no Anexo 6.

Os resultados mostrados na tabela 20 indicam que as variáveis gênero e faixa etária interagem quando analisamos o apagamento da vogal seguinte ao [l] (p-valor<0,001). Isso indica que a análise dessas variáveis deve levar em consideração a influência que uma exerce sobre o efeito da outra. Tal influência será melhor analisada nas seções seguintes.

6.2.2. Velarização da lateral – [l] ~ [ɫ]

A tabela abaixo apresenta um recorte do resultado da regressão tendo como variável dependente [l] ~ [ɫ], com inclusão de todas as variáveis independentes, selecionadas por stepwise, desconsiderando a possibilidade de interação:

Tabela 21: Saída do SPSS, com variável dependente [l] ~ [ɫ] e variáveis independentes gênero e faixa etária²⁸.

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	21 / 451	4,7	<0,001	0,29
	Masculino*	100 / 663	15,1		1,00
Faixa Etária	Adulto	35 / 512	6,8	0,009	0,55
	Jovem*	602 / 86	14,3		1,00

* fator de referência

A saída acima indica que gênero e faixa etária têm influência significativa na velarização da lateral. A inclusão do termo de interação entre *gênero e faixa etária* na variável [l] ~ [ɫ] não apresentou significância estatística, como pode ser visto no Anexo 7. Isso indica que a interação entre tais variáveis independentes não é estatisticamente significativa e que, portanto, o modelo final desta variável dependente não deve incluir o *termo de interação*.

6.2.3. Apagamento da sílaba - [l]+vogal ~ Ø

A tabela abaixo apresenta um recorte do resultado da regressão tendo como variável dependente [l]+vogal ~ Ø, com inclusão de todas as variáveis independentes, selecionadas por stepwise, desconsiderando a possibilidade de interação:

²⁸ Saída completa do SPSS no Anexo 7

Tabela 22: Saída do SPSS, com variável dependente [1]+vogal ~ Ø e variáveis independentes gênero e faixa etária²⁹

Variáveis Independentes	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	621 / 1336	46,5	<0,001	0,39
	Masculino*	588 / 944	62,3		1,00
Faixa Etária	Adulto	532 / 1012	52,6	0,008	0,76
	Jovem*	677 / 1268	53,4		1,00

* fator de referência

O resultado acima indica que as variáveis gênero e faixa etária apresentaram efeitos estatisticamente significativos.

A tabela abaixo apresenta o resultado da regressão tendo como variável dependente [1]+vogal e Ø, com inclusão de todas as variáveis independentes, selecionadas por stepwise, e inclusão do termo da interação entre gênero e faixa etária:

Tabela 23: Saída do SPSS, com variável dependente [1]+vogal ~ Ø e variáveis independentes gênero, faixa etária e o termo da interação entre gênero e faixa etária³⁰

Variáveis Independentes	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	621 / 1336	46,5	<0,001	0,25
	Masculino*	588 / 944	62,3		1,00
Faixa Etária	Adulto	532 / 1012	52,6	<0,001	0,43
	Jovem*	677 / 1268	53,4		1,00
Termo da Interação	Gênero _F x Faixa Etária _A			<0,001	2,53

* fator de referência

Os resultados mostrados na tabela 23 indicam que as variáveis gênero e faixa etária interagem quando analisamos o apagamento da sílaba (p-valor<0,001). Isso indica que a análise dessas variáveis deve considerar a influência que uma variável independente exercem sobre o efeito da outra. Tal influência será melhor analisada nas seções seguintes.

6.3. Seleção do modelo

Como já tratado na seção 5.3.1 a seleção dos modelos para cada variável dependente em análise será feita a partir do método *stepwise (forward-lr)*. Além das variáveis

²⁹ Saída completa do SPSS no Anexo 8.

independentes analisadas neste estudo, considerarei também os *termos da interação* estatisticamente significativos, conforme identificados na seção anterior. Seguem os modelos finais, a partir dos quais serão feitas as análises.

6.3.1. Variável [l]+vogal ~ Ø

Tabela 24: Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ Ø, no estilo entrevista, com termo da interação³¹

Variáveis Independentes	Fatores	n ₁ / n _r	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	621 / 1336	46,5	<0,001	0,25
	Masculino*	588 / 944	62,3		1,00
Faixa Etária	Adulto	532 / 1012	52,6	<0,001	0,43
	Jovem*	677 / 1268	53,4		1,00
Termo da Interação	Gênero _F x Faixa Etária _A			<0,001	2,53
Contexto Seguinte	Pausa	55 / 339	16,2	<0,001	0,11
	Consoante*	1041 / 1644	63,3	<0,001	1,00
	Vogal	113 / 297	38,0	<0,001	0,34
Classe da Palavra	Nome*	46 / 335	13,7	<0,001	1,00
	Pronome	1147 / 1855	61,8	<0,001	6,57
	Verbo	16 / 90	17,8	0,048	2,03
Classe da Palavra Seguinte	Verbo auxiliar*	110 / 143	76,9	<0,001	1,00
	Verbo não-aux.	657 / 965	68,1	<0,001	0,43
	Não-verbo	372 / 814	45,7	<0,001	0,33
	Pausa	70 / 358	19,6	0,051	0,42
Vogal na variável	Vogal A	329 / 909	36,2	0,546	0,82
	Vogal U*	17 / 115	14,8	<0,001	1,00
	Vogal I	863 / 1256	68,7	0,029	2,08
Tonicidade	Proparoxítona	13 / 23	56,5	<0,001	11,6
	Paroxítona*	1196 / 2257	53,0		1,00
/S/ final da palavra	Ausente*	877 / 1816	48,3	<0,001	1,00
	Presente	332 / 464	71,6		1,81

* fator de referência

³⁰ Saída completa do SPSS no Anexo 8.

³¹ Saída completa do SPSS no Anexo 8.

6.3.2. Variável [l]+vogal ~ [l]

Tabela 25: Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [l]+vogal ~ [l], no estilo entrevista, com termo da interação³².

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	430 / 1145	37,6	<0,001	0,21
	Masculino*	563 / 919	61,3		1,00
Faixa Etária	Adulto	477 / 957	49,8	0,001	0,62
	Jovem*	516 / 1107	46,6		1,00
Termo da Interação	Gênero _F x Faixa Etária _A			<0,001	2,62
Contexto Seguinte	Pausa	191 / 475	40,2	<0,001	0,28
	Consoante*	486 / 1089	44,6	<0,001	1,00
	Vogal	316 / 500	63,2	<0,001	2,04
Classe da Palavra Seguinte	Verbo auxiliar*	41 / 74	55,4	0,006	1,00
	Verbo não-aux.	325 / 633	51,3	0,381	0,79
	Não-verbo	413 / 855	48,3	0,392	0,79
	Pausa	214 / 502	42,6	0,037	2,35
Vogal na variável	Vogal A	257 / 837	30,7	<0,001	0,25
	Vogal U*	161 / 259	62,2	<0,001	1,00
	Vogal I	575 / 968	59,4	0,651	1,01
/S/ final da palavra	Ausente*	942 / 1881	50,1	<0,001	1,00
	Presente	51 / 183	27,9		0,28

* fator de referência

³² Saída completa do SPSS no Anexo 6.

6.3.3. Variável [1] ~ [ɿ]

Tabela 26: Modelo de regressão contendo todas as variáveis independentes selecionadas por stepwise para a variável dependente [1] ~ [ɿ], no estilo entrevista³³

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Gênero	Feminino	21 / 451	4,7	<0,001	0,29
	Masculino*	100 / 663	15,1		1,00
Faixa Etária	Adulto	35 / 512	6,8	0,009	0,55
	Jovem*	602 / 86	14,3		1,00
Contexto Seguinte	Pausa	73 / 264	27,7	0,031	4,53
	Consoante*	24 / 510	4,7	0,086	1,00
	Vogal	24 / 340	7,1	0,394	1,30
Classe da Palavra	Nome*	26 / 247	10,5	<0,001	1,00
	Pronome	93 / 817	11,4	<0,001	2,93
	Verbo	2 / 50	4,0	0,509	0,60
Classe da Palavra Seguinte	Verbo auxiliar*	1 / 42	2,4	0,023	1,00
	Verbo não-aux.	14 / 339	4,1	0,590	1,77
	Não-verbo	34 / 447	7,6	0,141	4,60
	Pausa	72 / 286	25,2	0,202	4,86
Tonicidade	Proparoxítona	3 / 6	50,0	0,005	18,4
	Paroxítona*	118 / 1108	10,6		1,00

* fator de referência

6.4. Variáveis sociais

6.4.1. Gênero e faixa etária

Como mostrado nas saídas para os modelos de regressão, na seção 6.3, as variáveis gênero e faixa etária interagem nas variáveis dependentes [1]_{+vogal} ~ \emptyset e [1]_{+vogal} ~ [1]. Isso indica que a análise de tais variáveis independentes deve ser feita de forma conjunta, já que uma variável interfere no efeito da outra variável nas variáveis dependentes.

³³ Saída completa do SPSS no Anexo 7.

Neste estudo, a análise da interação entre as variáveis gênero e faixa etária é fundamental na medida em que possibilita que os dados sejam analisados considerando que a variável gênero interfere no efeito da variável faixa etária na variável dependente, e vice-versa. A análise das variáveis independentes em interação, entretanto, não é igual à análise de variáveis independentes sem interação. Seus efeitos devem ser analisados de forma conjunta com o efeito do termo da interação. Devido a uma maior complexidade para a realização de tal análise, não a farei aqui, e opto por um agrupamento das variáveis independentes em interação. Assim, ao invés de incluir no modelo uma variável gênero, composta de dois fatores, uma variável faixa etária, composta de dois fatores, e o termo da interação entre tais variáveis, será considerado somente uma variável *gênero-faixa etária*, composta de quatro fatores: mulheres-jovens, mulheres-adultas, homens-jovens e homens-adultos. Essa decisão leva-nos a perdas em relação à análise das variáveis gênero e faixa etária, já que os seus efeitos no modelo de regressão não poderão ser observados separadamente; entretanto, permite-nos analisar de forma mais simples a interferência conjunta de tais variáveis. Apesar da interação entre gênero e faixa etária no fenômeno de velarização da lateral não apresentar-se como estatisticamente significativa, a análise de tal fenômeno também será feita a partir do agrupamento entre gênero e faixa etária, de forma que o resultado da variável dependente [l] ~ [ɫ] possa ser melhor comparado com os resultados para as outras variáveis dependentes.

Segue abaixo a tabela para a variável [l]_{+vogal} ~ ∅:

Tabela 27: Resultados para variável *gênero-faixa etária* na variável dependente [l]_{+vogal} ~ ∅, no estilo entrevista³⁴

[l] _{+vogal} ~ ∅	Fator	n ₁ / n _t	% ₁	p-valor	Razão de Chances
	masculino-jovem*	369 / 544	67,8		1,00
	masculino-adulto	219 / 400	54,8	<0,001	0,43
	feminino-jovem	308 / 724	42,5	<0,001	0,25
	feminino-adulto	313 / 612	51,1	<0,001	0,27

* fator de referência

A tabela acima indica que a variante ∅ é bastante favorecida pelo grupo masculino-jovem, tanto em termos percentuais, quanto em termos de razão de chances. A chance de um indivíduo do grupo masculino-jovem utilizar a variante ∅ é 4 vezes a chance de um

³⁴ Saída completa do SPSS no Anexo 11

indivíduo do grupo feminino-jovem. Entre os homens, a chance de um jovem utilizar a variante \emptyset é 2,3 vezes a chance de um adulto (1,0/0,43). Isso indica que os homens, de um modo geral, apagam mais a sílaba final do que as mulheres. Entre as mulheres, a diferença entre jovens e adultas é bem mais reduzida. Se colocarmos o grupo *feminino-adulto* como categoria de referência teremos um p-valor=0,568 para o grupo *feminino-jovem*, o que indica que a diferença entre efeitos dos fatores feminino-jovem e feminino adulto não é estatisticamente significativa. Colocando o grupo *feminino-jovem* como categoria de referência temos que um p-valor=0,004 para o grupo *masculino-jovem*. Esse resultado indica que não há diferença entre os efeitos dos grupos *feminino-jovem* e *feminino-adulto* e que, portanto, a faixa etária é mais importante para explicarmos a variabilidade entre os homens do que entre as mulheres.

Tabela 28: Resultados para variável *gênero-faixa etária* na variável dependente [1]+vogal ~ [1], no estilo entrevista³⁵

[1]+vogal ~ [1]	Fator	n ₁ / n _t	% ₀₁	p-valor	Razão de Chances
	masculino-jovem*	316 / 491	64,4		1,00
	masculino-adulto	247 / 428	57,7	0,001	0,62
	feminino-jovem	200 / 616	32,5	<0,001	0,21
	feminino-adulto	230 / 529	43,5	<0,001	0,34

* fator de referência

A tabela acima indica que a variável *gênero/faixa etária* apresenta alto poder explicativo sobre a implementação da variante [1], com diferenças importantes entre todos os grupos formados e um apagamento maior da vogal entre os homens, semelhante ao ocorrido no apagamento da sílaba. Para esta variável dependente temos que, entre os jovens, homens e mulheres ocupam posições opostas em relação à utilização da variante [1], sendo que, entre as mulheres jovens, a chance de utilização de tal variante é 0,21 vezes a chances do grupo formado por homens jovens. Entre as mulheres, as adultas têm 1,6 vezes a chances de utilizar a variante [1] em relação às jovens (0,34/0,21). Colocando o grupo feminino-adulto como referência, temos um p-valor<0,001 para o grupo feminino-jovem. Colocando o grupo feminino-jovem como referência, temos um p-valor<0,001 para o grupo masculino-jovem, o que indica que a diferença entre os efeitos entre todos os grupos é estatisticamente significativa. Entre os adultos, a chance de um homem utilizar a variante [1] é 1,8 vezes a chance de uma mulher (0,62/0,34).

³⁵ Saída completa do SPSS no Anexo 9.

A tabela abaixo apresenta os resultados da variável [l] ~ [ɫ]:

Tabela 29: Resultados para variável *gênero-faixa etária* na variável dependente [l] ~ [ɫ], no estilo entrevista³⁶

[l] ~ [ɫ]	Fator	n _l / n _t	% _l	p-valor	Razão de Chances
	masculino-jovem*	75 / 391	19,2		1,00
	masculino-adulto	25 / 272	9,2	0,003	0,45
	feminino-jovem	11 / 211	5,2	<0,001	0,21
	feminino-adulto	10 / 240	4,2	<0,001	0,22

* fator de referência

Os resultados apresentados na tabela acima são bastante semelhantes aos resultados dos processos apresentados anteriormente. Temos que a variante [ɫ] é muito favorecida pelo grupo masculino-jovem. A razão de chances entre as faixas etárias para as mulheres é a mesma, isso indica que o fenômeno de velarização é mais característico dos homens e, principalmente, dos homens jovens. Lembramos aqui que não houve interação entre gênero e faixa etária no processo de velarização. Assim, podemos dizer também que homens e jovens, separadamente, favorecem tal processo.

Os resultados apresentados para os processos em análise indicam que, em Itaúna, os homens utilizam mais as variantes mais inovadoras do que as mulheres. Pesquisas em sociolinguística variacionista têm mostrado que as mulheres estão à frente em alguns processos de mudança linguística. Em Labov (1994), o autor relaciona tais processos como aqueles que se encontram abaixo do nível de consciência social do fenômeno. Assim, nos fenômenos não perceptíveis pelos falantes, as mulheres frequentemente são líderes na mudança. Entretanto, em alguns fenômenos, as mulheres são mais conservadoras. Tais fenômenos seriam aqueles que estariam acima do nível de consciência social e que, além disso, a variante mais inovadora fosse estigmatizada. Com relação ao comportamento das faixas etárias, parece não haver regularidades ou generalizações, podemos dizer apenas que, de modo geral, homens jovens estão à frente dos processos. A variante velarizada insere-se como uma possível marcação social do grupo formado por falantes jovens do gênero masculino. A observação dos resultados entre os homens, mostra que os jovens são aqueles que utilizam mais as variantes mais inovadoras nos três processos em análise. O comportamento das mulheres parece não se diferenciar tanto, já que em dois processos, no apagamento da sílaba e na velarização, as diferenças entre jovens e adultas é muito pequena.

O fenômeno de apagamento da vogal apresentou diferença significativa entre os efeitos das mulheres jovens e das mulheres adultas. Isso poderia ser um indício de progressão do processo. Tal resultado pode indicar diferenças nos comportamentos individuais dos falantes analisados ou a possibilidade de outros agrupamentos sociais estarem atuando. Esse aspecto será retomado na análise do fator *indivíduo*.

6.4.2. Estilo

A variável *estilo* não foi inserida no modelo de regressão, já que os dados analisados correspondiam somente ao estilo *entrevista*. Como veremos nesta seção, tal variável fornece-nos informações interessantes com relação ao direcionamento da variabilidade em relação à gradações diferentes em termos de formalidade nos estilos considerados neste estudo. A tabela abaixo apresenta o cruzamento entre as variantes e o estilo, incluindo todos os dados coletados nesta pesquisa.

Tabela 30: Tabela de contingência entre a variável *estilo* e todas as variantes

Variantes		Entrevista	Perguntas Direcionadas	Leitura de Texto	Leitura de Palavras	Total
[l]+vogal	n	1071	108	1419	941	3539
	%	31,6	55,4	73,1	86,6	53,5
[l]	n	993	82	463	143	1681
	%	29,3	42,1	23,9	13,7	25,4
[ɫ]	n	121	4	18	2	145
	%	3,6	2,1	0,9	0,2	2,2
∅	n	1209	1	40	1	1251
	%	35,6	0,5	2,1	0,1	18,9
total	n	3394	195	1940	1087	6616
	%	100,0	100,0	100,0	100,0	100,0

$$\chi^2 = 1869,299, df=9, p\text{-valor}<0,001$$

Os resultados apresentados na tabela 30 sugerem que a variação em Itaúna é bastante sensível ao estilo, já que, de um modo geral, estilos mais formais favorecem variantes mais conservadoras.

A pouca ocorrência das variantes ∅ e [ɫ] nos estilos *perguntas direcionadas* e *leitura de palavras* pode indicar que tais fenômenos diferenciam-se do [l], especialmente em relação

³⁶ Saída completa do SPSS no Anexo 10.

a produções isoladas. Em tais estilos, os itens lexicais foram produzidos pelos falantes isoladamente, sem contextos adjacentes. Pode indicar também que na ausência da vogal seguinte à lateral a redução é menor e mais geral do que quando a sílaba final é apagada ou quando há velarização. Dizer [a'kɛl], para o item *aquele*, deve permitir que o sentido do item lexical seja mais previsível do que dizer [a'kɛ]. Quando o item lexical encontra-se em contexto, a porcentagem de apagamento da sílaba final é aumentada, pois o contexto poderia estar ajudando na atribuição do sentido.

A relação de dependência entre as variantes pode ser visualizada na tabela 30. A porcentagem de ocorrência da variante [l] em relação às outras variantes foi maior no estilo *perguntas direcionadas* do que no estilo *entrevistas*, o que, a princípio, seria um resultado inesperado. Entretanto, se observarmos as ocorrências da variante \emptyset em tais estilos veremos que o fenômeno de apagamento da sílaba final atua consideravelmente no estilo *entrevista*, o que fez com que a ocorrência da variante [l] fosse percentualmente menor, neste estilo.

A tabela abaixo apresenta as variantes [l]+vogal e [l] distribuídas entre os grupos sociais (gênero e faixa etária agrupados) nos estilos *entrevista* e *leitura de texto*, ambos com uma quantidade grande de dados e com itens lexicais produzidos em contexto, diferenciando-se somente pelo grau de formalidade. Os demais processos não serão analisados devido à baixa quantidade de dados no estilo *leitura de texto*.

Tabela 31: Tabela de contingência entre os fatores *entrevista* e *leitura de textos*, da variável *estilo* e a variável *gênero/faixa etária*, para a variável dependente [l]+vogal e [l].

Estilo		Grupos				Total	
		masculino jovem	masculino adulto	feminino jovem	feminino adulto		
entrevista	[l]+vogal	n	175	181	416	299	1071
		%	35,6	42,3	67,5	56,5	51,9
	[l]	n	316	247	200	230	993
		%	64,4	57,7	32,5	43,5	48,1
	Total	n	491	428	616	529	2064
		%	100,0	100,0	100,0	100,0	100,0
leitura de texto	[l]+vogal	n	296	411	327	385	1419
		%	65,2	86,9	68,7	80,4	75,4
	[l]	n	158	62	149	94	463
		%	34,8	13,1	31,3	19,6	24,6
	Total	n	454	473	476	479	1882
		%	100,0	100,0	100,0	100,0	100,0

entrevista: $\chi^2 = 132,650$, $df=3$, $p\text{-valor}<0,001$ / leitura de texto: $\chi^2 = 77,072$, $df=3$, $p\text{-valor}<0,001$

Na tabela acima, analisando percentuais, temos que o comportamento dos grupos sociais parece apresentar direções diferentes dependendo do estilo. No estilo entrevista, por exemplo, se observarmos as variantes [1]_{+vogal} e [1], parece haver um uso maior da variante [1] entre os homens e da variantes [1]_{+vogal} entre as mulheres. No estilo leitura de texto a idade parece ser mais importante, já que a porcentagem entre os jovens do gênero feminino e do gênero masculino são semelhantes no uso da variante [1]. Da mesma forma, homens adultos e mulheres adultas também comportam-se de maneira semelhante em relação à variável [1]_{+vogal} e [1], sendo as mulheres o grupo que mais utiliza a variante [1]. Tal constatação sugere que, na fala mais informal (estilo *entrevista*), os homens utilizam mais a variante mais inovadora [1].

6.5. Variáveis lingüísticas

6.5.1. Contexto seguinte

Tabela 32: Resultados do efeito da variável *contexto seguinte* na variável dependente [1]_{+vogal} ~ Ø

	Fatores	n_1 / n_t	% ₁	p-valor (Wald)	Razão de chances
Contexto Seguinte	Pausa	55 / 339	16,2	<0,001	0,11
	Consoante*	1041 / 1644	63,3		1,00
	Vogal	113 / 297	38,0	<0,001	0,34

* fator de referência

O resultados apresentados acima indicam que a chance de apagamento da sílaba final /l/_{+vogal} antes de consoante é 3 vezes a chance de apagamento antes de vogal e 9 vezes antes de pausa. Indica ainda que a chance de apagamento antes de vogal é 3 vezes a chance antes de pausa.

Os resultados obtidos na tabela 32 podem estar sendo muito influenciados pelos pronomes, conforme pode ser visto na tabela abaixo, em que os itens pronominais são excluídos da análise:

Tabela 33: Resultados do efeito da variável *contexto seguinte* na variável dependente [l]+vogal ~ Ø, excluídos os pronomes³⁷

	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
Contexto Seguinte	Pausa	17 / 124	12,1	0,029	0,43
	Consoante*	31 / 216	14,4		1,00
	Vogal	14 / 68	20,6	0,593	1,22

* fator de referência

Neste caso, o contexto seguinte *pausa* é desfavorecedor, conforme já foi visto na análise dos estilos. Os ambientes seguintes *consoante* e *vogal* não apresentaram efeitos estatisticamente diferentes, quando os dados relativos aos pronomes são excluídos. Esse resultado pode indicar que, entre os pronomes, algum fator relacionado ao contexto seguinte *consoante* pode estar atuando. Neste caso, seria necessária uma análise item a item em contextos para averiguar este resultado.

Tabela 34: Resultados do efeito da variável *contexto seguinte* na variável dependente [l]+vogal ~ [l]

	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
Contexto Seguinte	Pausa	191 / 475	40,2	<0,001	0,28
	Consoante*	486 / 1089	44,6		1,00
	Vogal	316 / 500	63,2	<0,001	2,04

* fator de referência

Os resultados acima indicam que a chance de ocorrer a variante [l] antes de vogal é 2 vezes a chance de ocorrer antes de consoante e 7,3 vezes a chance de ocorrer antes de pausa. Pode-se atribuir o efeito do contexto seguinte *vogal* a um processo de ressilabação vocálica ocorrida no português devido à elisão da vogal da variável, conforme Bisol (2000, p. 116), o que leva, por exemplo, a ocorrências do tipo [ɛ'leɾə] para “ela era” (IM15). Quando somente os itens que apresentam /S/ na sílaba /l/+vogal são analisados, a variável *contexto seguinte* não apresenta significância estatística, como pode ser visto no anexo 18. Isso indica que os resultados apresentados na tabela 34 referem-se principalmente aos itens que não apresentam /S/ e que, assim, pode-se atribuir o favorecimento da *vogal seguinte* a um processo de elisão. De forma semelhante ao que ocorre no apagamento da sílaba, no fenômeno de apagamento da vogal parece ocorrer uma oposição entre pausa e presença de segmento. Tal processo poderia ser caracterizado como um processo de redução motivado por efeitos de coarticulação, conforme Sproat e Fujimura (1993) e Bybee (2001).

³⁷ Saída completa do SPSS no Anexo 18.

Tabela 35: Resultados do efeito da variável *contexto seguinte* na variável dependente [l] ~ [ɫ]

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Contexto	Pausa	73 / 264	27,7	0,031	4,53
Seguinte	Consoante*	24 / 510	4,7		1,00
	Vogal	24 / 340	7,1	0,394	1,30

* fator de referência

O p-valor obtido pelo teste de Wald parece indicar que a variável *contexto seguinte* não apresenta efeitos estatisticamente significativos entre os seus fatores. Entretanto, a variável *contexto seguinte* foi selecionada como significativa pelo procedimento stepwise, que utiliza como critério o teste da razão da máxima verossimilhança. Isso indica que um modelo de regressão para a variável dependente [l] ~ [ɫ] que contenha a variável independente *contexto seguinte* é melhor do que um modelo que não contenha tal variável independente. O efeito gerado para tal variável em termos de razão de chances deve ser observado com cautela. Se observarmos na tabela abaixo as ocorrências da variante velarizada em itens não-pronominais, veremos que a ocorrência desta variante está praticamente restrita ao contexto seguinte *pausa*. Esse resultado coloca o fenômeno de velarização em oposição aos fenômenos de apagamento da sílaba e de apagamento da vogal, em relação ao favorecimento do contexto-seguinte.

Tabela 36: Cruzamento da variável dependente [l] ~ [ɫ] com a variável *contexto seguinte*, sem pronomes.

		Consoante	Vogal	Pausa	Total
[l]	n	111	81	77	269
	%	97,4	100,0	75,5	90,6
[ɫ]	n	3	0	25	28
	%	2,6	0,0	24,5	9,4
Total	n	114	81	102	297
	%	100,0	100,0	100,0	100,0

6.5.2. Contexto anterior

A variável contexto anterior não se mostrou estatisticamente significativa para nenhum dos processos analisados, como pode ser visto nos Anexos 6, 7 e 8, no qual são expostos os modelos completos, com a inserção de todas as variáveis analisadas neste estudo. Isso indica que a variação nos itens terminados em sílaba átona /l/+vogal parece não ser influenciada pela vogal anterior.

6.5.3. Classe da palavra

Tabela 37: Resultados do efeito da variável *classe da palavra* na variável dependente [l]+vogal ~ Ø

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Classe da Palavra	Nome*	46 / 335	13,7		1,00
	Pronome	1147 / 1855	61,8	<0,001	6,57
	Verbo	16 / 90	17,8	0,048	2,03

* fator de referência

Os resultados apresentados na tabela 37 indicam que a variante Ø é mais utilizada quando o item é um pronome, do que quando não é. As diferenças percentuais revelam que a diferença entre o uso de Ø entre pronomes e não-pronomes é bastante grande, o que poderia indicar que o fenômeno de apagamento da sílaba final /l/+vogal é um fenômeno que atinge principalmente os pronomes e que talvez esteja alastrando-se para as outras classes. Isso pode ser observado a partir das ocorrências apresentadas na tabela 37 para a classe dos *nomes* (46 em 335) e para a classe dos *verbos* (16 em 90). Entre os verbos, em que houve apagamento da sílaba, 13 em 16 referem-se aos itens *falo* (9) e *fala* (4). Entre os nomes, em que houve apagamento da sílaba, 18 em 46 referem-se aos itens *divinópolis* (8) e *escola* (10). Os demais *nomes*, nos casos em que ocorreu apagamento da sílaba, referem-se a 18 itens. Isso indica que os resultados apresentados são bastante influenciados por uma quantidade pequena de itens que apresentam apagamento da sílaba.

O fato do apagamento em itens não-pronominais restringir-se a poucos itens levamos a dois questionamentos: 1. por que os pronomes estão mais sujeitos ao apagamento da sílaba do que não-pronomes? e 2. por que somente alguns itens não-pronominais apresentaram ocorrências de Ø? Uma proposta para estas questões poderia relacionar-se à frequência de uso dos itens, caso o fenômeno seja caracterizado como motivado foneticamente. Na seção 6.6.3 apresentarei alguns resultados relacionados à interferência da frequência na variabilidade em Itaúna.

A tabela abaixo apresenta o resultado da regressão logística com os itens lexicais *ele/eles/ela/elas* excluídos:

Tabela 38: Resultados do efeito da variável *classe da palavra* na variável dependente [l]+vogal ~ Ø excluídos os pronomes *ele/eles/ela/elas*

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Classe da Palavra	Nome*	46 / 335	13,7		1,00
	Pronome	170 / 162	51,2	<0,001	3,23
	Verbo	16 / 90	17,8	0,194	0,77

* fator de referência

A tabela acima mostra que outros pronomes (*aquele(s), aquela(s), nele(s), nela(s)*, entre outros) também apresentam mais apagamento da sílaba final do que nas demais classes. No entanto, o efeito gerado com a inclusão dos pronomes *ele/eles/ela/elas* é bem maior do que sem tais itens (6,6 para 3,2), o que pode indicar que algum outro fator sintático ou semântico não mensurado pode estar atuando no processo de apagamento da sílaba.

A variável independente *classe da palavra* não apresentou significância estatística para a variável dependente [l]+vogal ~ [l] o que indica que, nos dados coletados em Itaúna, não podemos associar nenhum efeito em tal variável dependente à classe da palavra, conforme Anexo 6. Esse resultado indica que este fenômeno é mais abrangente do que o apagamento da sílaba.

A tabela abaixo apresenta os resultados para a variável *classe da palavra* na fenômeno de velarização da lateral:

Tabela 39: Resultados do efeito da variável *classe da palavra* na variável dependente [l] ~ [ɫ]

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Classe da Palavra	Nome*	26 / 247	10,5		1,00
	Pronome	93 / 817	11,4	<0,001	2,93
	Verbo	2 / 50	4,0	0,509	0,60

* fator de referência

O resultado acima indica que, assim como no apagamento da sílaba /l/+vogal, o fenômeno de velarização ocorre mais em itens pronominais do que em nomes.

6.5.4. Vogal da variável

Tabela 40: Resultados do efeito da variável *vogal da variável* na variável dependente [1]+vogal ~ Ø

	Fatores	n_1 / n_t	% ₀₁	p-valor (Wald)	Razão de chances
Vogal na variável	Vogal A	329 / 909	36,2	0,546	0,82
	Vogal U*	17 / 115	14,8		1,00
	Vogal I	863 / 1256	68,7	0,029	2,08

* fator de referência

Na tabela 40, o p-valor para a *vogal A* indica que o seu efeito não se diferencia estatisticamente do efeito do fator de referência *vogal U*. Vemos que a vogal *i* seguinte ao /l/ favorece o apagamento da sílaba /l/+vogal mais do que as demais vogais. Tal constatação pode ser explicada se considerarmos os resultados apresentados para a variável dependente [1]+vogal ~ Ø na seção anterior, ou seja, a classe pronominal favorece o apagamento da sílaba. Entre os pronomes, 2/3 correspondem a pronomes terminados em *i*, como pode ser visto na tabela abaixo:

Tabela 41: Distribuição da variável *vogal da variável* entre pronomes na variável dependente [1]+vogal ~ Ø

Vogal	Ocorrências	%
u	10	0,5
i	1209	65,2
a	636	34,3
Total	1855	100,0

Como a quantidade de pronomes é muito maior do que a quantidade de itens não-pronominais e a quantidade de pronomes terminados em *i* é muito maior do que terminados em *a* ou em *u* temos que o resultado apresentado na tabela 40 é um reflexo da má distribuição dos dados e seria preciso fazer outros refinamentos para uma análise mais precisa. Ao que parece, existe uma questão fonética que atua neste processo, favorecidos pela vogal *i*, mas existem também outros aspectos que não puderam ser analisados neste trabalho.

Tabela 42: Resultados do efeito da variável *vogal da variável* na variável dependente [1]+vogal ~ [1]

	Fatores	n_1 / n_t	% ₀₁	p-valor (Wald)	Razão de chances
Vogal na variável	Vogal A	257 / 837	30,7	<0,001	0,25
	Vogal U*	161 / 259	62,2		1,00
	Vogal I	575 / 968	59,4	0,651	1,01

* fator de referência

Os resultados apresentados na tabela 42 indicam que o apagamento da vogal está relacionado à sua altura, já que a chance de apagamento da vogal para as vogais *i e u* correspondem a quase quatro vezes a chance de apagamento para a vogal *a*.

No processo de velarização da lateral a vogal posterior ao /l/ não apresentou influência estatisticamente significativa, conforme Anexo 7.

6.5.5. Classe da palavra seguinte

Tabela 43: Resultados do efeito da variável *classe da palavra seguinte* na variável dependente [l]+vogal ~ [Ø]

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Classe da	Verbo auxiliar*	110 / 143	76,9		1,00
Palavra Seguinte	Verbo não-aux.	657 / 965	68,1	<0,001	0,43
	Não-verbo	372 / 814	45,7	<0,001	0,33
	Pausa	70 / 358	19,6	0,051	0,42

* fator de referência

A variável classe da palavra seguinte apresentou resultados interessantes para a variável dependente [l]+vogal ~ Ø. A tabela 43 indica que a chance de ocorrer apagamento da sílaba antes de um verbo auxiliar é três vezes a chance de ocorrer antes de não-verbo e mais de duas vezes a chance de ocorrer antes de verbos não-auxiliares. A explicação para tal diferença parece não poder ser dada pela hipótese apresentada em Bybee (2001, p.186). Segundo a autora, estruturas compostas por pronome+auxiliar, por serem agrupamentos muito freqüentes, têm maior tendência a reduções fonéticas.

Os resultados apresentados na tabela 41 correspondem a todos os itens analisados, conforme anexo 4; entretanto, 67% (1.527 em 2.280) dos itens correspondem aos pronomes ele, elas, eles e elas e, por esse motivo, é provável que o resultado apresentado esteja sendo influenciado por tais itens. Refazendo a análise, somente com os itens ele/ela/eles/elas, com todas as variáveis independentes incluídas e seleção pelo método forward, temos a seguinte tabela para a variável *classe da palavra seguinte*:

Tabela 44: Resultados do efeito da variável *classe da palavra seguinte* na variável dependente [l]+vogal ~ Ø, somente para os itens ele/ela/eles/elas

[l]+vogal ~ Ø	Fator	n ₁ / n _t	% ₁	p-valor	Razão de Chances
Classe da palavra Seguinte	Verbo Auxiliar*	104 / 133	78,2		1,00
	Verbo não-Auxiliar	631 / 905	69,7	0,003	0,47
	Não-verbo	215 / 354	60,7	< 0,01	0,32
	Pausa	30 / 135	22,2	0,022	0,29

* fator de referência

A tabela acima mostra que a presença de um verbo auxiliar seguinte aos pronomes ele/ela/eles/elas afeta significativamente a realização da variável [l]+vogal ~ Ø, sendo a variante Ø favorecida pela presença de um verbo auxiliar seguinte. A diferença entre as demais classes não é significativa quando alteramos a categoria de referência para *pausa*. Neste caso, temos os seguintes p-valores: verbo auxiliar: p-valor=0,022; verbo não-auxiliar: p-valor=0,304; não-verbo: p-valor=0,776 e pausa*: p-valor<0,001. Se colocarmos o fator *não-verbo* como fator de referência, teremos um p-valor=0,018 para o fator *verbo não-auxiliar*, o que indica que a diferença entre os efeitos de tais fatores é estatisticamente significativa. Isso quer dizer que o *verbo não-auxiliar* também favorece ligeiramente o processo de apagamento da sílaba.

Os resultados parecem indicar que tais itens de fato estão mais sujeitos à variação quando são seguidos de verbo auxiliar. A presença de um verbo auxiliar seguinte não significa que o pronome esteja na posição de sujeito. Neste estudo, a função sintática não foi controlada. Entretanto, uma observação do banco de dados permite dizer que, na grande maioria dos casos, pronomes antes de verbos têm função de sujeito. Dessa forma, o resultado apresentado acima pode apresentar indícios de que estruturas pronome+auxiliar estão mais propensas à realização da variante Ø, não porque sejam mais freqüentes, mas por algum outro fator possa estar atuando.

A hipótese apresentada em Correa (1998) é de que a motivação para o processo variável nestes itens é de ordem sintática, já que, como argumenta o autor, o fenômeno estaria relacionado a um processo de cliticização, ocasionado pelo uso do pronome *eles* como indeterminador do sujeito. Nas seções seguintes, analisarei a interferência da freqüência de ocorrência dos itens nos processos variáveis em análise (seção 6.6.3) e a interferência dos *eles* como indeterminador do sujeito (seção 6.5.6).

A tabela abaixo refere-se à interferência da classe seguinte na variável [l]+vogal ~ [l]

Tabela 45: Resultados do efeito da variável *classe da palavra seguinte* na variável dependente [l]+vogal ~ [l]

	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
Classe da Palavra Seguinte	Verbo auxiliar*	41 / 74	55,4		1,00
	Verbo não-aux.	325 / 633	51,3	0,381	0,79
	Não-verbo	413 / 855	48,3	0,392	0,79
	Pausa	214 / 502	42,6	0,037	2,35

* fator de referência

O resultado mostrado acima indica que as classes *verbo auxiliar*, *verbo não-auxiliar* e *não-verbo* não apresentam diferenças significativas entre seus efeitos. Assim, o único fator que apresentou diferença em relação aos demais fatores foi o fator *pausa*, que na realidade não é uma classe. Se excluirmos os dados de *pausa* do modelo, a variável classe da palavra seguinte não é selecionada pelo método stepwise. O resultado apresentado na tabela parece estranho se compararmos com os resultados da variável *contexto seguinte*, que atribuiu à *pausa* o menor efeito no apagamento da vogal. Talvez o efeito atribuído à *pausa*, neste caso, possa estar sendo causado por alguma interação com outra variável.

A variável *classe da palavra seguinte* não apresentou significância para o fenômeno de velarização da lateral.

6.5.6. Fricativa alveolar seguinte

Tabela 46: Resultados do efeito da variável *fricativa alveolar seguinte* na variável dependente [l]+vogal ~ Ø

	Fatores	n ₁ / n _t	% ₀₁	p-valor (Wald)	Razão de chances
/S/ final da palavra	Ausente*	877 / 1816	48,3		1,00
	Presente	332 / 464	71,6	<0,001	1,81

* fator de referência

O resultado apresentado acima indica que a presença de /S/ seguinte à sílaba final /l/+vogal, como em *aqueles*, favorece o apagamento da sílaba átona final formada por /l/+vogal. A chance de ocorrer a variante Ø com a presença do /S/ seguinte corresponde a 1,8 vezes a chance de ocorrer tal variante quando o /S/ seguinte não está presente. Isso indica que a chance de um indivíduo falar *aquês* para *aqueles* é bem maior do que a chance de um indivíduo falar *aquê* para *aquele*.

Tal constatação já havia sido feita em Correa (1998) para o pronome *eles*. Segundo o autor, a explicação para tal fato está no uso do pronome *eles* como indeterminador do sujeito e, conseqüentemente, a um processo de cliticização de tal pronome na língua portuguesa.

Os dados coletados em Itaúna indicam que o processo de apagamento da sílaba final /l/+vogal seguida de /S/ não está restrito ao item *eles*, como pode ser visto na tabela abaixo, e que, portanto, a hipótese de que o item *eles* favorece o apagamento da sílaba devido a seu uso como indeterminador do sujeito parece não ser o único fator que influencia o apagamento da sílaba:

Tabela 47: Ocorrências de itens que apresentam /S/ final e apagamento da sílaba final /l/+vogal

	[l]+vogal		Ø	
	n	%	n	%
aquelas	3	21,43	12	80,00
aqueles	2	12,50	14	87,50
daquelas	0	0,00	2	100,00
daqueles	1	33,33	2	66,67
deles	8	32,00	17	68,00
Divinópolis	6	46,15	7	53,85
elas	4	23,53	13	76,47
eles	74	21,80	266	78,20
óculos	0	0,00	3	100,00

A partir das ocorrências do item *eles* nos dados coletados em Itaúna, criei uma nova variável independente para testar se o fato do pronome *eles* ser indeterminado favorece que a sílaba final /l/+vogal seja apagada. Assim, a partir de um total de 341 dados, todos do pronome *eles*, incluindo todas as variáveis independentes, obtive o seguinte resultado para a variável *indeterminação*:

Tabela 48: Resultados do efeito da variável *indeterminação* na variável dependente [l]+vogal ~ Ø para o item *eles*

Fatores		n_1 / n_t	% ₁	p-valor (Wald)	Razão de chances
Indeterminação	Determinado*	72 / 95	75,8	0,846	1,00
	Indeterminado	196 / 246	79,7		1,10

* fator de referência

Os resultados da tabela 48 indicam que, nos dados de Itaúna, o efeito do pronome *eles* utilizado como indeterminador do sujeito não apresentou diferenças estatisticamente significativas em relação às situações em que o *eles* atua como um sujeito determinado.

Uma proposta para explicar o favorecimento do /S/ no apagamento da sílaba pode estar relacionada à quantidade de material fonético nos casos em que ocorre /S/. O falante utilizaria a variante \emptyset em itens como *ês*, para *eles*, mais do que em *ê*, para *ele*, porque a presença do /S/ possibilitaria que o sentido do item reduzido fosse mais facilmente retomado pelo ouvinte. Isso pode ser observado se analisarmos os dados apresentados na tabela abaixo, em que podem ser comparadas as frequências de apagamento de itens com /S/ e itens sem /S/ em outros pronomes.

Tabela 49: Comparação entre itens com /S/ e itens sem /S/ na variável [l]+vogal ~ \emptyset

	[l]+vogal		\emptyset	
	n	%	n	%
aquela	26	53,1	23	46,9
aquelas	3	20,0	12	80,0
aquele	7	24,1	22	75,9
aqueles	2	12,5	14	87,5
daquela	0	0,00	3	100,00
daquelas	0	0,00	2	100,00
daquele	2	40,0	3	60,0
daqueles	1	33,3	2	66,7
dele	41	51,9	38	48,1
deles	8	32,0	17	68,0
ela	255	54,0	217	46,0
elas	4	23,5	13	76,5
ele	213	30,7	481	69,3
eles	74	21,8	266	78,2

Outra proposta, seria assumir o *princípio de licenciamento da coda*, conforme Kaye (1989). Segundo tal princípio, as consoantes em coda são seguidas de um núcleo vazio, não preenchido. Em tal princípio, segundo Cristófaró-Silva (1999, p. 213), a posição de *coda* “deve ser seguida por uma posição de *onset* que a governa”. Neste caso, itens paroxítonos terminados em /S/ seriam proparoxítonos, sendo a sílaba final formada por /S/+núcleo vazio. Assim, no caso da ocorrência de *ês* para *eles*, teríamos um apagamento da sílaba medial, semelhante ao que ocorre nas proparoxítonas, como em [ˈɔkus], para *óculos* ou [diviˈnɔps], para *Divinópolis*. Sobre esse aspecto, precisaríamos de um estudo mais aprofundado em relação às proparoxítonas.

Ao contrário do que ocorre no apagamento da sílaba, no apagamento da vogal da sílaba /l/+vogal a presença do /S/ seguinte inibe a ocorrência da variante [l], como pode ser visto na tabela abaixo.

Tabela 50: Resultados do efeito da variável *fricativa alveolar seguinte* na variável dependente [l]+vogal ~ [ɫ]

Fatores		n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
/S/ final da palavra	Ausente*	942 / 1881	50,1		1,00
	Presente	51 / 183	27,9	<0,001	0,28

* fator de referência

O resultado acima poderia ser explicado em termos da fonotática da língua portuguesa. Bybee (2001, p. 13) afirma que diversos estudos têm mostrado que a fonotática de uma língua é estabelecida em termos probabilísticos, ou seja, pela frequência de uso de padrões sonoros. O padrão /l+/S/, na mesma sílaba, parece tratar-se de um padrão pouco frequente na língua portuguesa, entretanto, esse assunto merece uma investigação mais profunda, já que foram encontradas ocorrências de /l+/S/ na fala de Itaúna, como pode ser visto na tabela 50.

A presença do /S/ seguinte não apresentou significância estatística para o fenômeno de velarização da lateral, portanto, tal variável independente parece não interferir no processo de velarização. Isso ocorre porque os dados desta variável incluem somente os casos em que a vogal seguinte ao /l/ foi apagada, ou seja, as ocorrências das variantes [l] e [ɫ]. Assim, o desfavorecimento da presença do /S/ para a variante [l] também causa um desfavorecimento para a variante [ɫ].

6.5.7. Tonicidade

As tabelas abaixo apresentam os resultados para a variável *tonicidade* nas variáveis dependentes [l]+vogal ~ Ø e [l] ~ [ɫ]:

Tabela 51: Resultados do efeito da variável *tonicidade* na variável dependente [l]+vogal ~ Ø

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Tonicidade	Proparoxítona	13 / 23	56,5	<0,001	11,6
	Paroxítona*	1196 / 2257	53,0		1,00

* fator de referência

Tabela 52: Resultados do efeito da variável *tonicidade* na variável dependente [l] ~ [ɫ]

Variáveis Independentes	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Tonicidade	Proparoxítona	3 / 6	50,0	0,005	18,4
	Paroxítona*	118 / 1108	10,6		

* fator de referência

Os resultados apresentados nas tabelas 51 e 52 indicam que a tonicidade interfere significativamente na realização das variantes \emptyset e [ɫ], nos processos de apagamento da sílaba /l/+vogal e velarização da lateral, respectivamente. Entretanto, tal constatação precisa ser feita com ressalvas, dadas as distribuições dos dados nas tabelas. Como pode ser visto, a ocorrência de palavras proparoxítonas é bastante reduzida. Na variável [l]_{+vogal} ~ \emptyset , como pode ser visto na tabela 51, a diferença de 3,5% entre paroxítonas e proparoxítonas parece não ser tão significativa, dada a diferença na quantidade de dados em tais fatores (2257 e 23, respectivamente). Da mesma forma, a diferença de porcentagem entre paroxítonas e proparoxítonas no fenômeno de velarização também precisa ser vista com ressalvas, dada a pequena ocorrência de proparoxítonas nesses dados (6). A tabela abaixo apresenta todas as ocorrências de proparoxítonas, no estilo entrevista, distribuídas entre as variantes:

Tabela 53: Distribuição das proparoxítonas entre as variantes no estilo entrevista

Itens	[l] _{+vogal}	[l]	[ɫ]	\emptyset	Total
ângulo	0	0	2	0	2
Divinópolis	6	1	0	8	15
fórmula	1	0	0	1	2
Mariângela	0	1	0	0	1
matrícula	0	0	0	1	1
óculos	0	0	0	3	3
ridículo	1	0	0	0	1
título	1	0	0	0	1
triângulo	0	1	0	0	1
Total	9	3	2	13	27

Na tabela acima, vemos que na variável [l]_{+vogal} ~ \emptyset , 64% dos itens proparoxítonos correspondem ao item *Divinópolis* ([l]_{+vogal}=6 e \emptyset =8). As demais ocorrências correspondem aos itens *fórmula* ([l]_{+vogal}=1 e \emptyset =1), *matrícula* ([l]_{+vogal}=1 e \emptyset =1), *óculos* ([l]_{+vogal}=0 e \emptyset =3), *ridículo* e *título* ([l]_{+vogal}=1 e \emptyset =0). Isso indica que o resultado da regressão, na tabela 51, é muito influenciado pelo item lexical *Divinópolis* e que talvez não possamos fazer generalizações.

6.6. Outras variáveis

6.6.1. Indivíduo

A análise dos resultados para as variáveis sociais *gênero* e *faixa etária*, apresentados na seção 6.4.1, fornece informações importantes relacionadas à variabilidade entre os grupos sociais na cidade de Itaúna. Entretanto, somente as características sociais analisadas parecem não ser suficientes para explicar a variabilidade. Tais resultados indicam que os homens utilizam mais as variantes mais inovadoras do que as mulheres; os homens jovens, em especial, formam o grupo que mais favorece o uso das variantes mais inovadoras. As mulheres jovens, no fenômeno de apagamento da vogal, apresentaram-se como um grupo mais conservador. Nos demais processos, jovens e adultas apresentam efeitos semelhantes. O fato da faixa etária influenciar homens e mulheres de maneira diferente sugere a necessidade de averiguarmos mais profundamente os comportamentos individuais dos indivíduos pesquisados.

A análise do comportamento individual tem sido proposta em diversos estudos que consideram a variação como objeto de estudo. Em Paiva e Duarte (2003), apresentam-se estudos sobre o comportamento do indivíduo a partir do estudo da mudança lingüística em tempo real na fala do Rio de Janeiro. Em Oliveira (1997), o autor defende que o comportamento do indivíduo não é uma reprodução do grupo ao qual ele pertence. Segundo o autor, indivíduos de um mesmo grupo social podem apresentar comportamentos completamente diferentes. Uma análise pormenorizada do comportamento individual pode trazer algumas reflexões a respeito da variabilidade lingüística na comunidade de fala analisada.

Para a análise dos indivíduos, utilizarei um método estatístico chamado Classificação e Regressão em Árvores, ou CART. Tal método foi utilizado, em sociolingüística, no estudo feito por Mendoza-Denton, Hay e Jannedy (2003).

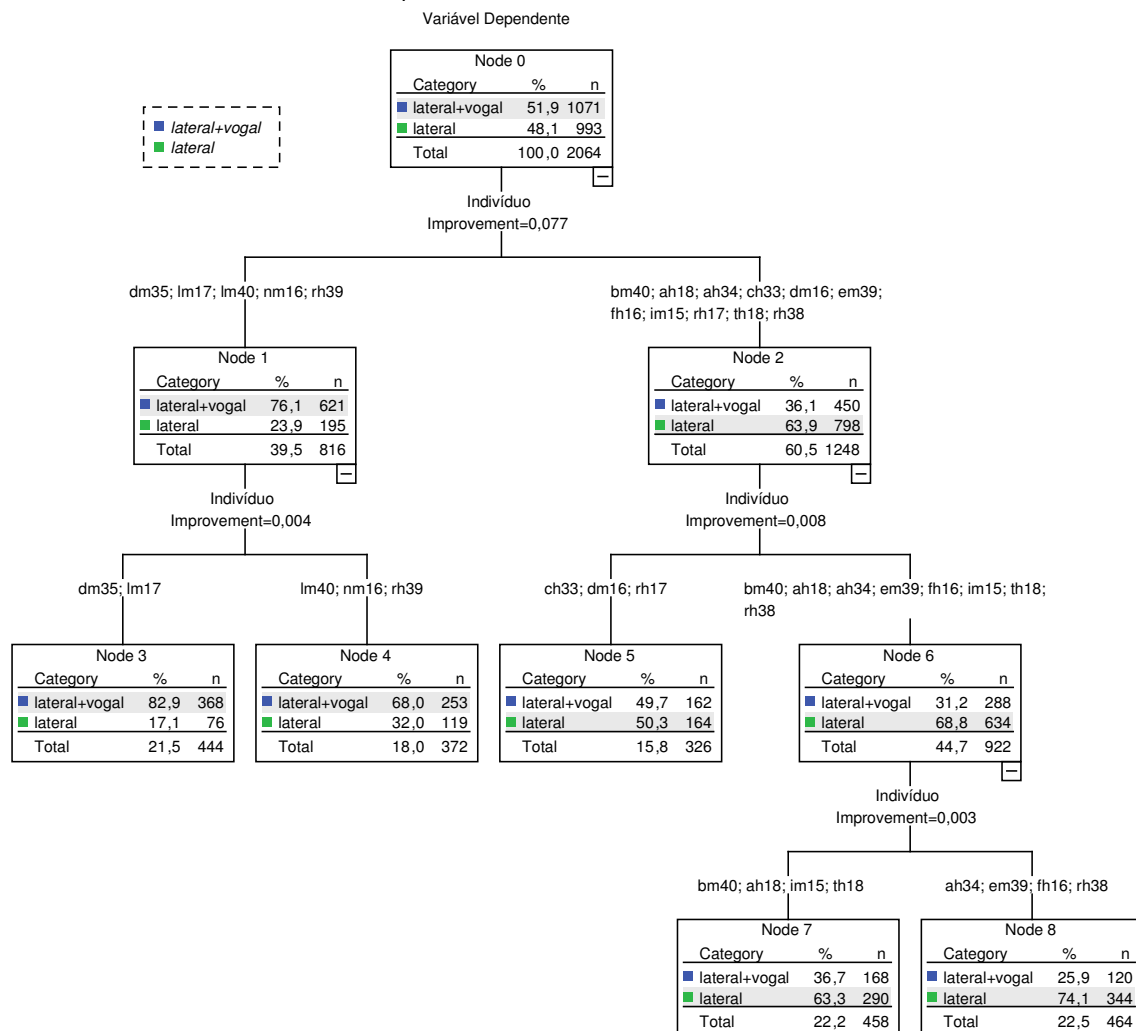
Segundo Ferreira, Soares e Cruz (2001), o modelo utilizado no CART, “é ajustado mediante sucessivas divisões binárias no conjunto de dados, de modo a tornar os subconjuntos resultantes cada vez mais homogêneos, em relação à variável resposta”. Assim, a utilização de tal método permite-nos reagrupar os indivíduos da amostra em grupos homogêneos em relação à variável dependente em análise. Os novos agrupamentos

serão feitos desconsiderando a estratificação social feita a priori, entre gênero e faixa etária, e considerando somente o comportamento dos indivíduos em relação ao uso das variantes.

Os códigos atribuídos aos indivíduos permitem identificar o gênero e a faixa etária ao qual eles pertencem. As letras *b* e *m* referem-se, respectivamente, a *homem* e *mulher*. Os números referem-se à idade do informante, classificados como jovens, de 15 a 20, e adultos, de 30 a 40 anos.

A árvore abaixo refere-se ao CART para a variável $[l]_{+vocal} \sim [l]$, tendo como variável independente o indivíduo:

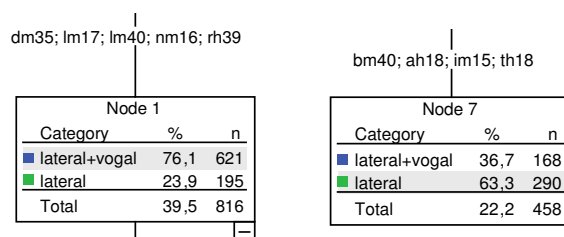
Figura 11: CART para a variável dependente $[l]_{+vocal} \sim [l]$ e variável independente *indivíduo*, no estilo *entrevista*



A árvore acima mostra que os agrupamentos criados para a variável $[l]_{+vocal} \sim [l]$ não seguem todos os agrupamentos entre as variáveis sociais gênero e faixa etária consideradas nesta pesquisa. A partir dele, vemos que as variáveis gênero e faixa etária não

são suficientes para explicarmos a variabilidade encontrada nos dados. Tomemos, por exemplo, os nós 1 e 7:

Figura 12: Recorte da Figura 11



O agrupamento formado pelo nó 1 mostra-se bastante heterogêneo quanto à faixa etária dos indivíduos, porém mais homogêneo com relação ao gênero. A partir dele podemos sugerir que o gênero feminino é mais conservador no uso da variante mais inovadora [1] e que a faixa etária não é muito determinante. O agrupamento formado pelo nó 7, ao contrário, é mais heterogêneo com relação ao gênero e mais homogêneo com relação à faixa etária (jovens usando mais).

A heterogeneidade dos nós sugere que, para que pudéssemos explicar a variabilidade em termos de características sociais, necessitaríamos fazer um estudo mais amplo e mais profundo dos informantes entrevistados, para que, talvez, pudéssemos encontrar características comuns que os agrupassem de acordo com o agrupamento dos nós. Vale lembrar que todas as demais variáveis sociais, comumente consideradas nos estudos variacionistas, foram controladas na seleção dos informantes, conforme já descrito no capítulo 4, ou seja, grupo social, escolaridade, região e procedência dos pais. O objetivo era criar um grupo socialmente mais homogêneo, diferenciado somente pela faixa etária e pelo gênero. A criação de grupos socialmente homogêneos, entretanto, não é suficiente para determinarmos a homogeneidade dos estratos sociais. Pode ser que existam agrupamentos sociais atuando no processo, pois as redes sociais (Milroy, L., 1987) não foram controladas. Em Labov (2001), o autor ressalta a importância dos inovadores, aqueles que iniciam o processo, e dos líderes, aqueles responsáveis pelo espalhamento do processo, nos fenômenos de variação e mudança lingüística.

A partir dos agrupamentos de indivíduos criados pelo CART foi criada uma nova variável com 5 fatores, no qual os indivíduos alocados nos nós eram agrupados como pertencentes a um mesmo grupo. Utilizei a divisão do nó 6 entre os nós 7 e 8 para que a

quantidade de indivíduos nos grupos fosse semelhante. Os resultados podem ser vistos na tabela abaixo.

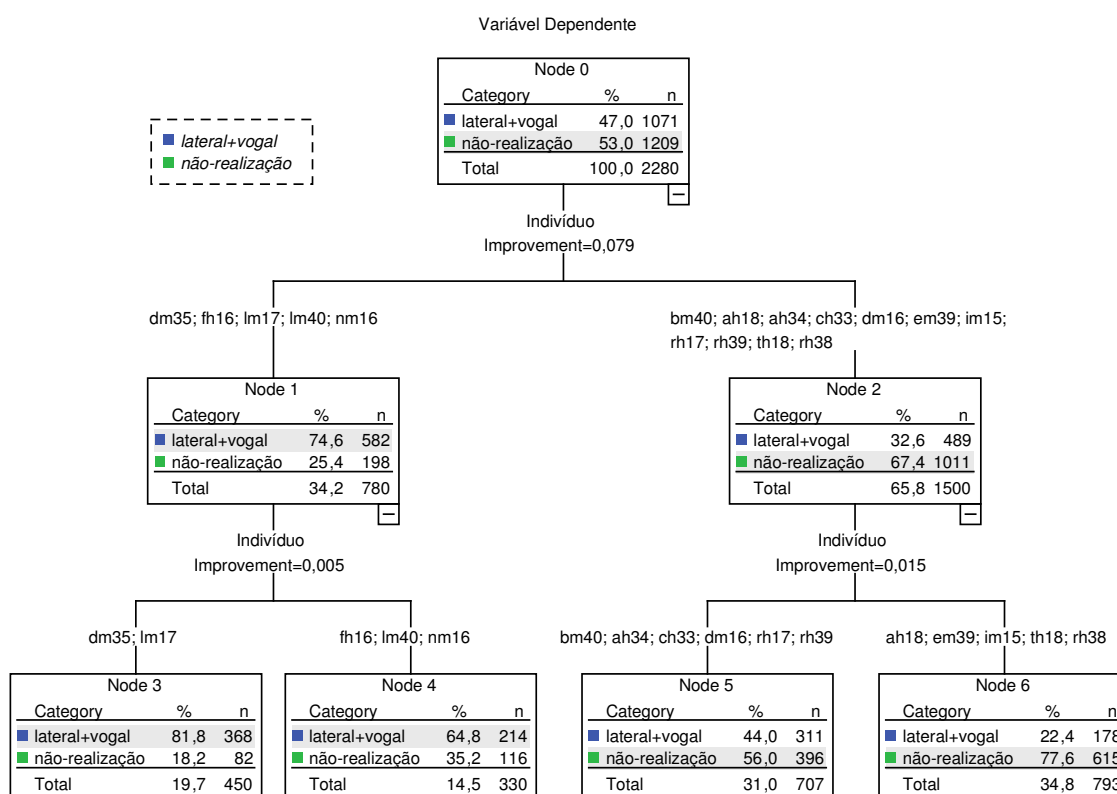
Tabela 55: Resultados do efeito do reagrupamento dos indivíduos na variável dependente [1]_{+vogal} ~ [1]

Grupos	Indivíduos	p-valor (Wald)	Razão de Chances
Nó 3*	dm35; lm17		1,0
Nó 4	lm40; nm16; rh39	<0,001	2,3
Nó 5	ch33; dm16; rh17	<0,001	4,9
Nó 7	bm40; ah18; im15; th18	<0,001	8,4
Nó 8	ah34; em39; fh16; rh38	<0,001	14,2

Os resultados acima indicam que os grupos criados são socialmente heterogêneos. Na seção 6.4.1 vimos que o grupo formado por mulheres jovens seria um grupo mais conservador em relação ao uso da variante [1], entretanto, a chance da informante im15 utilizar a variante [1] é oito vezes a chance do uso de tal variante pela informante lm17, ambas jovens do gênero feminino. A distribuição dos informantes entre os grupos mostra uma concentração maior de mulheres naqueles que têm menos chance de utilizar a variante [1] e de homens entre os grupos que têm mais chance de utilizar tal variante. Entretanto, nada impede que tal concentração tenha sido dada por características individuais ou por características sociais não mensuradas dos informantes da amostra selecionada, dada a heterogeneidade dos grupos formados.

A utilização do método CART nos dados para a variável [1]_{+vogal} ~ \emptyset também sugere que, apesar dos resultados da análise das variáveis gênero e faixa etária indicarem resultados interessantes e bastante explicativos do ponto de vista da variabilidade, os agrupamentos individuais com relação ao uso das variantes não seguem a estratificação em termos de variáveis sociais. Isso parece indicar que gênero e faixa etária são importantes para a variabilidade, mas também que o comportamento dos indivíduos da amostra não pode ser explicado somente por esses critérios, ao contrário, algum fator social ou individual não mensurado parece interferir na realização da variante \emptyset entre os indivíduos.

Figura 13: CART para a variável dependente [l]+vogal ~ Ø e variável independente *indivíduo*, no estilo *entrevista*

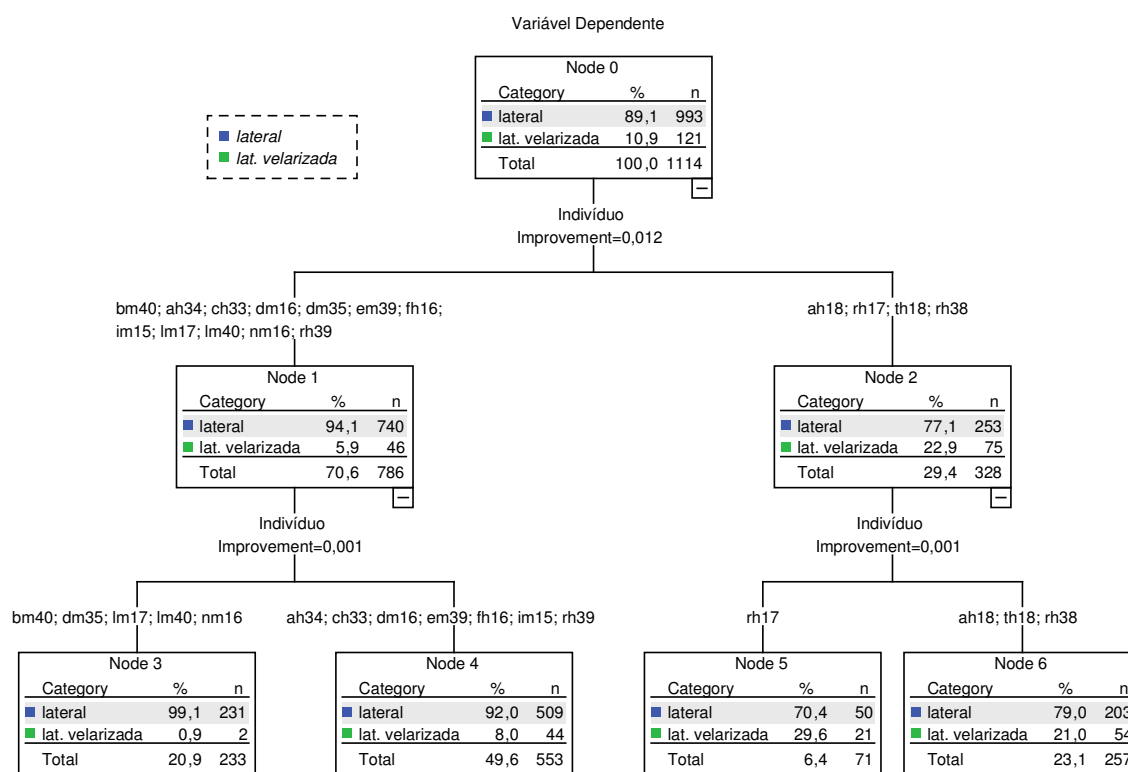


Ao compararmos as árvores das variáveis [l]+vogal ~ [l] e [l]+vogal ~ Ø, vemos que, em geral, os agrupamentos de indivíduos nos nós são os mesmos, com exceção de 4 informantes. Três deles mudaram para o nó seguinte (ah18 e rh39) ou para o nó anterior (bm40). A única mudança mais significativa foi do indivíduo fh16 que foi do último nó (6), ou seja, o grupo que mais utiliza a variante [l], para o 2º nó (4), que utiliza menos a variante Ø.

A pequena mobilidade dos indivíduos entre os nós nos dois processos reforça a idéia de que características não mensuradas neste estudo, atribuídas aos indivíduos, podem contribuir para que eles utilizem uma ou outra forma variante. Para um estudo mais detalhado da interferência de características sociais nos processos variáveis, talvez um estudo relacionado às redes sociais possa trazer algumas respostas, conforme L. Milroy (1987).

A análise da árvore para a variável [l] ~ [ɫ] parece reforçar a constatação de que a variante velarizada está relacionada a uma marcação do grupo formado por homens jovens:

Figura 14: CART para a variável dependente [l] ~ [ɮ] e variável independente *indivíduo*, no estilo *entrevista*



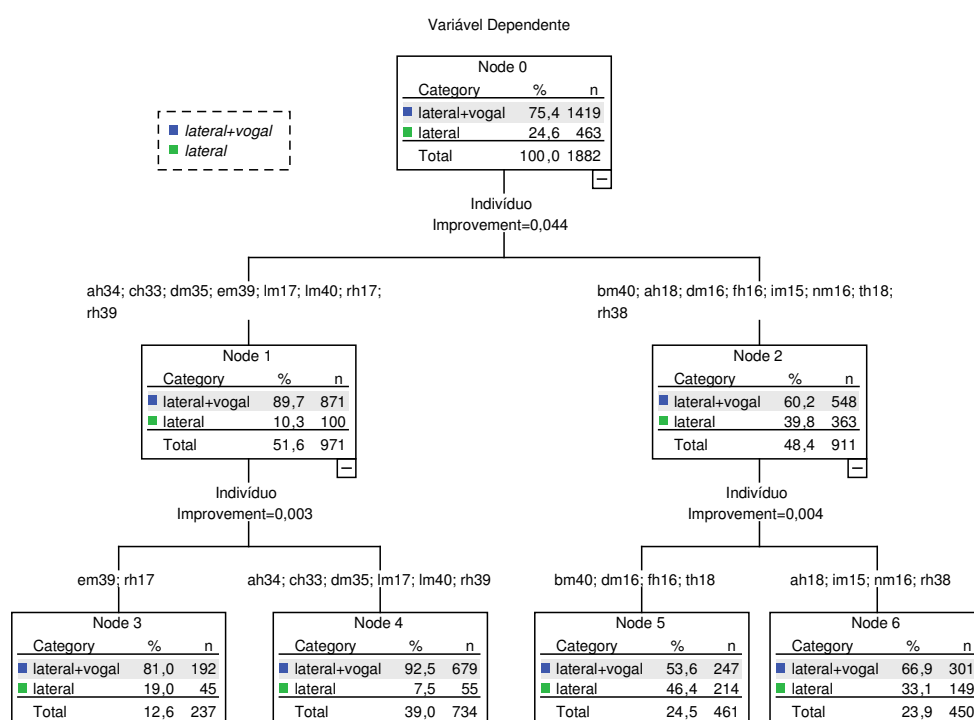
Como podemos ver, os grupos criados para o fenômeno de velarização da lateral apresentam uma homogeneidade maior do que para os demais fenômenos, tanto que não houve interação entre gênero e faixa etária nos dados. Neste caso, temos uma aglomeração dos indivíduos do gênero masculino como o grupo que utiliza mais a variante [ɮ], como no nó 2. Dos 4 informantes jovens do gênero masculino entrevistados, 3 deles encontram-se neste grupo. O nó 3, que agrupa os indivíduos que utilizam com menor frequência a variante [ɮ] é composto somente de mulheres, sendo que 3 das 4 mulheres adultas entrevistadas foram alocadas neste nó. O nó mais heterogêneo é o nó 4, que apresenta indivíduos de ambos os gêneros e ambas as faixas etárias, sendo 3 adultos do gênero masculino, 2 jovens do gênero feminino, 1 adulta do gênero feminino e 1 jovem do gênero masculino.

Uma questão que poderia ser levantada em relação à análise das árvores considerando-se somente o fator indivíduo é de que as informações relacionadas ao contexto no qual as observações foram produzidas não foram consideradas para a divisão dos nós. O modelo CART possibilita análises multivariadas e, inclusive, detecta

automaticamente as interações entre tais variáveis no banco de dados. Neste estudo, por questões de tempo, limitei a análise das árvores ao fator indivíduo. Entretanto, acredito que uma análise multivariada seja fundamental para compreendermos melhor o comportamento individual. Tal análise poderá ser feita em trabalhos futuros.

Uma forma de identificar se de fato a variabilidade entre os indivíduos decorre de características individuais ou sociais atribuídas a eles seria analisar o estilo *leitura de texto*. Neste estilo, todos os indivíduos pesquisados produziram os mesmos itens e em contextos lingüísticos idênticos. A única variável que permite esta análise é a variável dependente [1]+vogal ~ [1], as demais variáveis não apresentam dados suficientes no estilo *leitura de texto*. Segue abaixo a árvore para a variável [1]+vogal ~ [1], no estilo *leitura de texto*.

Figura 15: CART para a variável dependente [1]+vogal ~ [1] e variável independente *indivíduo*, no estilo *leitura de texto*



A árvore acima indica que, mesmo em contextos idênticos, os nós formados são relativamente heterogêneos. Entretanto, de um modo geral, parece haver uma divisão entre jovens e adultos no uso da variável [1] – índice de progressão do processo – apesar de certos falantes não se encaixarem nesta divisão, como por exemplo os falantes rh17 e bm40. Como pode ser visto acima, falantes jovens tendem a utilizar mais a variante [1] do que falantes adultos. Na análise dos estilos já havia sido mostrado que as diferenças entre o

comportamento entre jovens do gênero masculino e jovens do gênero feminino não são muito significativas na utilização da variante [l], no estilo *leitura de texto*.

6.6.2. Os itens lexicais

Nesta seção analisarei a realização dos itens lexicais distribuídos pelas variantes analisadas nesta dissertação. Tal análise justifica-se pela ênfase do estudo do item lexical na perspectiva dos modelos difusionistas, entre eles, a fonologia de uso.

Na tabela abaixo, apresentam-se todos os itens lexicais analisados neste estudo, no estilo *entrevista*, distribuídos em relação às 4 variantes para a sílaba final átona /l/+vogal:

Tabela 56: Itens lexicais distribuídos em relação às variantes da sílaba final átona /l/+vogal, no estilo *entrevista*

Item Lexical	[l]	[ɫ]	[ɮ]	∅	Total	% _{total}	% _{≠[l]+vogal}
amarela	1	0	0	0	1	0,03	0,00
ângulo	0	0	2	0	2	0,06	100,00
aquela	26	6	0	23	55	1,62	52,73
aquelas	3	2	0	12	17	0,50	82,35
aquele	7	16	1	22	46	1,36	84,78
aqueles	2	1	0	14	17	0,50	88,24
aquilo	4	24	1	2	31	0,91	87,10
asilo	25	15	6	2	48	1,41	47,92
asilos	2	2	0	0	4	0,12	50,00
aula	19	8	0	1	28	0,82	32,14
aulas	2	0	0	0	2	0,06	0,00
baile	2	0	0	0	2	0,06	0,00
bala	1	0	0	0	1	0,03	0,00
balas	1	1	0	0	2	0,06	50,00
belo	1	24	0	2	27	0,80	96,30
bola	14	17	2	1	34	1,00	58,82
bolo	5	4	0	0	9	0,27	44,44
cabelo	2	4	1	2	9	0,27	77,78
caçula	2	0	0	0	2	0,06	0,00
cavalo	1	0	0	0	1	0,03	0,00
chinchila	1	1	0	0	2	0,06	50,00
chinelos	0	0	1	0	1	0,03	100,00
circula	0	0	0	1	1	0,03	100,00
clientela	1	0	0	0	1	0,03	0,00
cola	3	4	0	0	7	0,21	57,14
colo	0	2	0	0	2	0,06	100,00
controle	2	0	0	0	2	0,06	0,00
costela	1	0	0	0	1	0,03	0,00
cotovelo	0	2	0	0	2	0,06	100,00
daquela	0	1	0	3	4	0,12	100,00
daquelas	0	0	0	2	2	0,06	100,00
daquele	2	2	0	3	7	0,21	71,43

Item Lexical	[l]	[l]	[ɫ]	Ø	Total	% _{total}	% _{≠[l]+vogal}
daqueles	1	1	0	2	4	0,12	75,00
daquilo	0	2	0	0	2	0,06	100,00
dela	32	31	13	24	100	2,95	68,00
delas	0	1	0	1	2	0,06	100,00
dele	41	91	25	38	195	5,75	78,97
deles	8	5	0	17	30	0,88	73,33
desfila	1	0	2	0	3	0,09	66,67
desfile	6	4	0	3	13	0,38	53,85
desfiles	3	0	0	0	3	0,09	0,00
desfilo	1	0	0	0	1	0,03	0,00
Divinópolis	6	1	0	8	15	0,44	60,00
ela	256	91	10	218	575	16,94	55,48
elas	4	1	0	13	18	0,53	77,78
ele	213	386	32	481	1112	32,76	80,85
eles	74	37	1	268	380	11,20	80,53
escola	55	33	3	10	101	2,98	45,54
escolas	13	2	0	0	15	0,44	13,33
estilo	0	3	0	0	3	0,09	100,00
fala	57	21	0	10	88	2,59	35,23
falo	8	25	0	4	37	1,09	78,38
favela	13	1	0	0	14	0,41	7,14
formula	1	0	0	1	2	0,06	50,00
gaiola	1	0	1	1	3	0,09	66,67
gaiolas	0	1	0	0	1	0,03	100,00
gola	1	0	0	0	1	0,03	0,00
intervalo	0	1	0	0	1	0,03	100,00
janela	0	0	1	0	1	0,03	100,00
lula	11	2	0	1	14	0,41	21,43
mala	0	5	0	2	7	0,21	100,00
malas	2	0	0	0	2	0,06	0,00
Mariângela	0	1	0	0	1	0,03	100,00
matrícula	0	0	0	1	1	0,03	100,00
Medioli	1	0	0	0	1	0,03	0,00
Melo	1	6	1	1	9	0,27	88,89
naquela	1	1	0	4	6	0,18	83,33
naquele	1	4	0	1	6	0,18	83,33
naquilo	0	2	0	0	2	0,06	100,00
nela	6	5	5	0	16	0,47	62,50
nele	9	11	5	1	26	0,77	65,38
novela	6	2	0	0	8	0,24	25,00
novelas	1	0	0	0	1	0,03	0,00
óculos	0	0	0	3	3	0,09	100,00
parcela	0	0	0	1	1	0,03	100,00
Paula	1	1	0	0	2	0,06	50,00
Paulo	5	6	0	0	11	0,32	54,55
pela	23	4	0	0	27	0,80	14,81
pelas	1	0	0	0	1	0,03	0,00
pele	0	1	0	0	1	0,03	100,00
pelo	28	19	0	0	47	1,38	40,43
pelos	4	0	0	0	4	0,12	0,00
pesadelo	0	1	0	0	1	0,03	100,00
pula	3	0	0	0	3	0,09	0,00

Item Lexical	[ɫ]	[l]	[ɣ]	∅	Total	% _{total}	%≠[ɫ]+vogal
pulo	0	2	0	0	2	0,06	100,00
quilos	1	0	0	0	1	0,03	0,00
ridículo	1	0	0	0	1	0,03	0,00
rola	2	1	0	0	3	0,09	33,33
rolo	0	1	0	0	1	0,03	100,00
sala	16	13	0	2	31	0,91	48,39
salas	3	0	0	0	3	0,09	0,00
tela	0	1	0	0	1	0,03	100,00
titulo	1	0	0	0	1	0,03	0,00
tranqüila	7	4	1	2	14	0,41	50,00
tranqüilo	6	24	7	1	38	1,12	84,21
triângulo	0	1	0	0	1	0,03	100,00
tutela	2	0	0	0	2	0,06	0,00
vela	3	1	0	0	4	0,12	25,00
total	1071	993	121	1209	3394	100,0	
%	31,6	29,3	3,6	35,6	100,0		

A tabela acima mostra que os itens lexicais parecem ser atingidos de modo diferenciado pelos processos variáveis. De um modo geral, como foi mostrado nas seções anteriores, os pronomes são os itens que apresentam a maior incidência da variante ∅, com raras exceções, como, por exemplo, os itens *fala* e *escola*. De maneira geral, para cada um dos pronomes, a maior quantidade de suas ocorrências quase sempre corresponde à variante ∅.

As ocorrências do item *escolas* vão contra a tendência dos pronomes de apagamento da sílaba final quando ela é seguida de /S/. Tal item depõe contra as duas possibilidades de explicação apresentadas para a constatação, a partir dos dados, de que o /S/ favorece o apagamento da sílaba. O item *escola* diferencia-se por apresentar uma ocorrência relativamente grande da variante ∅ (10/101), uma característica mais presente entre os itens pronominais. O mesmo ocorre com o item *fala* (10/88).

O item *aquilo* apresenta um comportamento diferenciado dos demais pronomes demonstrativos. Neste item, a ocorrência da variante ∅ (2/31) é muito menor do que a ocorrência da variante [ɫ] (24/31). O item *tranqüilo* porta-se de maneira semelhante, apresentando baixa ocorrência de ∅ (1/38) e alta ocorrência de [ɫ] (24/38). É interessante notar que o item *tranqüilo* é, muitas vezes, utilizado como cumprimento.

As ocorrências do item *belo* referem-se, na realidade, ao item *Belo Horizonte*. Para este caso, temos uma porcentagem de 89% de apagamento da vogal. De maneira

semelhante, o item *Melo* apresenta uma porcentagem elevada da variante (67%). Diferentemente do item *belo*, que apresenta uma motivação dada pelo contexto seguinte, o apagamento da vogal no item *Melo* parece estar relacionado a características específicas deste item. Todas as ocorrências do item *Melo* referem-se à escola José de Melo, bastante familiar, com o item contendo a variante [1] sendo pronunciado como [zɛdʒi'mɛl].

A porcentagem que cada item representa individualmente no estilo entrevista revela um dado interessante: 61% de todo o *corpus* analisado nesta pesquisa corresponde a 3 itens: ela (17%), ele (33%) e eles (11%).

As observações acima, retiradas da tabela 56, permitem-nos sugerir que os processos variáveis em Itaúna parecem atingir de maneira diferenciada os itens lexicais, mas é preciso fazer um estudo mais detalhado dos itens, dos contextos e das frequências de cada um para uma conclusão mais sólida. No decorrer deste trabalho foi possível identificar o fenômeno de apagamento da sílaba como um processo característico dos pronomes e presente já nos primórdios da língua portuguesa, como mostrado no capítulo 2. Como mostrado na tabela 56, este processo parece estar difundindo-se em direção a itens não-pronominais. O item *escola*, por exemplo, apresentou 10% de ocorrências para a variante \emptyset . A quantidade de vezes que o item ocorreu no total do *corpus* parece indicar que tal item apresenta alta frequência de ocorrência. Da mesma forma, observando os dados coletados nos jornais de Itaúna, temos que o item *escola* ocorreu 901 vezes num total de 2 milhões de palavras, o que caracteriza-o como item de alta frequência. Diante disso, é possível identificar indícios de que o espalhamento do processo de apagamento da sílaba final para itens não pronominais se dê por difusão lexical, e que esteja atingindo primeiramente e de forma gradual os itens mais frequentes. Acredito que esta é uma hipótese interessante a ser verificada futuramente.

6.6.3. A frequência

Em Bybee (2001), a autora afirma que as mudanças de etiologia fonética ocorrem primeiramente em itens de frequência de token mais alta do que em itens de frequência mais baixa. Conforme as características associadas aos processos de etiologia fonética em Bybee (op.cit., p. 65-69) podemos dizer que: (1) com relação ao grau de variabilidade entre as alternâncias, parece-nos que os processos relacionados à sílaba final átona /l/+vogal são foneticamente graduais. Essa suposição pode ser depreendida, principalmente, pela

identificação de sons intermediários entre as variantes, como vogais mais reduzidas, e pela dificuldade de categorização perceptual em alguns itens lexicais. (2) o domínio de aplicação dos processos vai além dos limites do item lexical, já que foram identificados processos de elisão vocálica entre o item com /l/+vogal final e o item seguinte, causando um processo de ressilabação. Nesta seção, apresentarei uma proposta de verificação da terceira característica, apresenta por Bybee para fenômenos de etiologia fonética: a relação entre variação e mudança lingüística e a frequência de *token*.

Para testar a interferência da frequência nos processos variáveis em Itaúna, utilizarei três *corpora*:

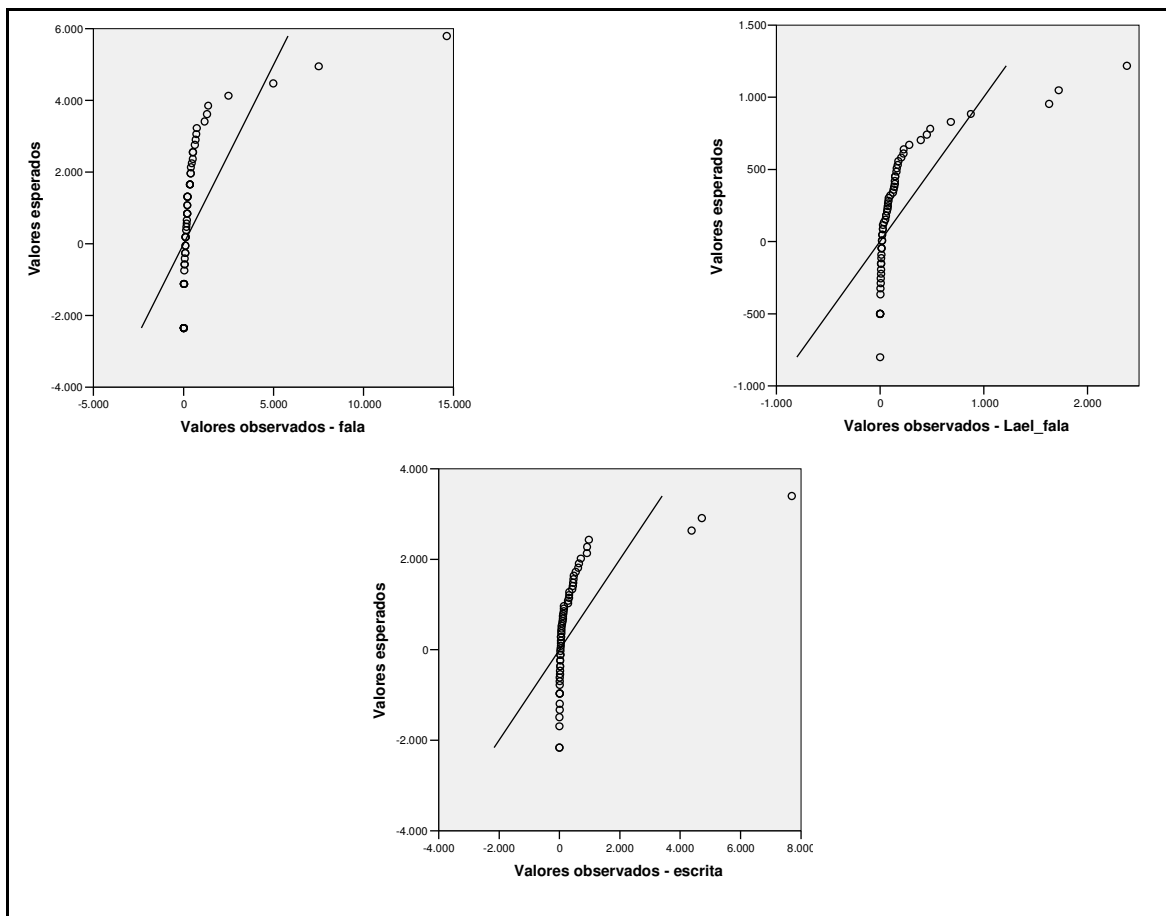
1. *Corpus* de fala, coletado nas entrevistas: este *corpus* seria o mais interessante, já que reflete com mais fidelidade a fala cotidiana de Itaúna como ela realmente ocorre, entretanto, a quantidade baixa de *tokens* (76.027 itens) traz limitações ao seu uso exclusivo;
2. *Corpus* de escrita, coletados dos jornais de Itaúna: este *corpus* apresenta uma grande quantidade de *tokens* (aproximadamente 2 milhões) e é interessante porque foi coletado na região em que a pesquisa foi feita, entretanto, o fato de ser um *corpus* de escrita faz com que ele não seja o ideal para ser utilizado na análise comparativa da fala.
3. *Corpus* de fala do LAEL-PUC/SP: o *corpus* de fala do Lael apresenta uma quantidade grande de ocorrências (aproximadamente 3 milhões) e foi coletado a partir da fala, entretanto, reflete a fala de uma região diferente da pesquisada neste estudo e, por isso, também não seria o *corpus* ideal para análise da fala de Itaúna.

As limitações impostas para o estudo com base na frequência do item, dadas pela ausência de *corpus* e pela dificuldade de se coletar dados numericamente significativos levamos a buscar alternativas para a análise da frequência de ocorrência. A alternativa que escolhi, para esta pesquisa, foi utilizar, ao mesmo tempo, os três conjuntos de dados de que disponho, cada um com a sua limitação.

O primeiro passo foi averiguar até que ponto os *corpora* utilizados poderiam ser de fato comparados. Para isso, criei um banco de dados em que inseri todos os itens analisados em Itaúna e associei tais itens à quantidade de vezes que eles apareciam nos três

conjuntos de dados utilizados, criando assim, 3 variáveis: *fala*, *Lael_fala* e *escrita*. A partir desse banco de dados, seria necessário identificar a correlação³⁸ entre tais variáveis. O estimador de correlação mais usual é chamado de *coeficiente de correlação de Pearson*, que assume que os dados têm *distribuição normal*. Nos gráficos abaixo, podemos ver que os dados nas variáveis *fala*, *Lael_fala* e *escrita* parecem não ter distribuição normal e que, portanto, devemos utilizar outro estimador, chamado *coeficiente de correlação de Spearman*.

Figura 16: Gráficos de distribuição dos dados para as variáveis *fala*, *Lael_fala* e *escrita* (gráfico Q-Q Plot, distribuição normal)



De acordo com Pagano e Gauvreau (2004), o coeficiente de *correlação de Pearson* é muito sensível a valores atípicos (aqueles que se dispersam), como os valores apresentados nos gráficos acima. O *coeficiente de correlação de Spearman* cria uma espécie de ranqueamento dos dados, minimizando o efeito dos valores atípicos. Este *coeficiente* assume valores entre -1 e 1. Valores próximos a -1 e 1 indicam alto grau de correlação entre as variáveis; valores próximos a 0 indicam que não há associação linear entre elas. A tabela abaixo apresenta os *coeficientes de correlação* para os três conjuntos de dados:

³⁸ A correlação quantifica o grau da relação linear entre as variáveis.

Tabela 57: Coeficientes de correlação de Spearman entre as variáveis fala, Lael_fala e escrita

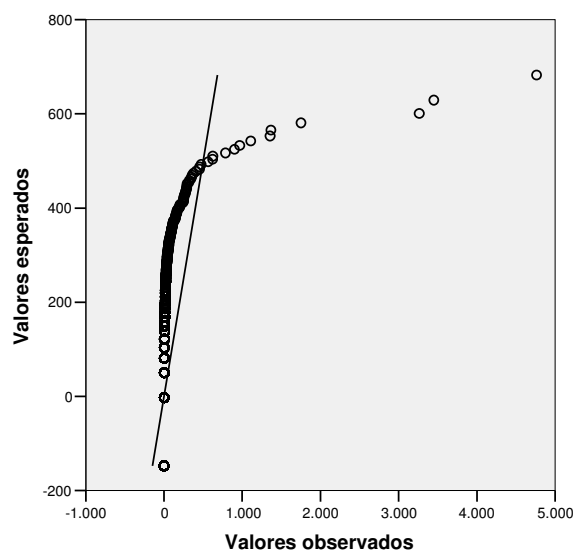
		Fala	Lael_fala	Escrita
Fala	Coeficiente	1,00	0,67	0,57
	p-valor	-	<0,001	<0,001
Lael_fala	Coeficiente	0,67	1,00	0,69
	p-valor	<0,001	-	<0,001
Escrita	Coeficiente	0,57	0,69	1,00
	p-valor	<0,001	<0,001	-

Os resultados obtidos na tabela acima indicam que a associação entre os bancos de dados de *fala*, do *Lael* e de *escrita* é estatisticamente significativa e que, portanto, os três conjuntos de dados são correlacionados.

Identificada a correlação entre os *corpora*, o passo seguinte seria definir os cortes nos bancos de dados. O que seriam considerados itens de baixa, média ou alta frequência? Quantas vezes um item teria que ocorrer para que ele fosse considerado de alta frequência?

Ao tentar fazer cortes com critérios estatísticos, baseados em valores médios, constatei que isso não seria interessante devido à distribuição desigual entre a frequência de ocorrência dos itens, ou seja, ocorreram poucos itens com frequência muito alta e muitos itens com frequência muito baixa, como pode ser visualizado no gráfico abaixo:

Figura 17: Gráfico de distribuição da frequência dos itens lexicais terminados em sílaba átona final /l/+vogal nos dados de fala de Itaúna (Q-Q Plot, distribuição normal)



Como a hipótese a ser testada é se a frequência do item interferia na sua variabilidade, associei a cada item a probabilidade de realização de formas variantes diferentes da forma plena [l]+vogal no *corpus* de fala de Itaúna, ou seja, aquelas formas que sofriam algum processo variável. Assim, no banco de dados que continha todos os itens analisados em Itaúna associados à quantidade de vezes que eles apareciam nos três conjuntos de dados, criei uma nova coluna que continha a probabilidade do item não ocorrer como [l]+vogal (Anexo 12).

A observação dos itens em relação à frequência de ocorrência e a probabilidade de não ocorrer nele a variante [l]+vogal possibilitou criar um corte, até certo ponto arbitrário, mas também motivado pela distribuição dos dados. Agrupei como itens mais frequentes aqueles que tinham frequência maior ou igual a 160 ocorrências por milhão e como itens de baixa frequência, aqueles que tinham frequência menor ou igual a 159 ocorrências por milhão.

Para testar interferência da variável independente *frequência de ocorrência do item* na variável dependente *probabilidade de não ocorrer [l]+vogal*, criei variáveis dicotômica, para que pudéssemos utilizar um modelo de regressão logística. Assim, estabeleci os seguintes critérios:

- ✓ Variável dependente:
 - 0 apagamento < 50%
 - 1 apagamento ≥ 50%

- ✓ Variável independente
 - 0 < 160 ocorrências / milhão
 - 1 ≥ 160 ocorrências / milhão

A partir de modelos de regressão logística para cada banco de dados, tendo como variável dependente a *probabilidade de não ocorrer [l]+vogal* e como variável independente a *frequência de ocorrência do item*, são obtidos os seguintes resultados

Tabela 58: Resultados do efeito da variável *freqüência de ocorrência do item* no banco de dados de fala de Itaúna na variável dependente *probabilidade de não ocorrer [l]+vogal*

ENTREVISTAS	Fator	n ₁ / n _t	% ₀₁	p-valor	Razão de Chances
Freqüência do item	< 160/mil	7 / 27	25,9	0,03	1,0
	≥ 160/mil	17 / 32	53,1		3,2

Tabela 59: Resultados do efeito da variável *freqüência de ocorrência do item vogal* no banco de dados de escrita de Itaúna na variável dependente *probabilidade de não ocorrer [l]+vogal*

JORNAIS	Fator	n ₁ / n _t	% ₀₁	p-valor	Razão de Chances
Freqüência do item	< 160/mil	17 / 45	37,8	0,418	1,0
	≥ 160/mil	7 / 14	50,0		1,6

Tabela 60: Resultados do efeito da variável *freqüência de ocorrência do item* no banco de dados do LAEL na variável dependente *probabilidade de não ocorrer [l]+vogal*

LAEL	Fator	n ₁ / n _t	% ₀₁	p-valor	Razão de Chances
Freqüência do item	< 160/mil	12 / 39	30,8	0,03	1,0
	≥ 160/mil	12 / 20	60,0		3,4

Os resultados apresentados acima indicam significância estatística para a análise dos bancos de dados de fala de Itaúna e do LAEL. O dados dos jornais de Itaúna não apresentaram significância (p-valor=0,418).

As razões de chances obtidas nas tabelas 58 e 60 (3,2 e 3,4) indicam que a chance de um item mais freqüente não ocorrer de forma plena em Itaúna é mais de 3 vezes a chance de um item menos freqüente. Esse resultado pode ser um indício de que há relação entre variação e freqüência do item.

Pudemos apresentar indícios que podem confirmar a hipótese levantada pela fonologia de uso, de que mudanças de etiologia articulatória ocorrem primeiramente em itens de freqüência de token mais alta do que em itens de freqüência mais baixa. Pudemos observar também que a seleção dos itens parece ser lexical, atingindo primeiramente os itens mais freqüentes.

6.7. Teste de avaliação das variantes

O objetivo do teste de avaliação das variantes aplicado aos falantes entrevistados na cidade de Itaúna foi tentar identificar se os informantes apresentavam algum tipo de consciência dos processos e, além disso se carregava algum estigma social. Tal teste foi realizado a partir dos procedimentos descritos na seção 4.4.3. Cada falante entrevistado avaliou 42 frases, listadas no Anexo 3.

Ao final do teste, perguntei a cada um dos informantes os critérios que eles utilizaram para atribuir níveis de escolaridade a partir da forma como uma pessoa fala. Algumas respostas a esta pergunta estão transcritas abaixo:

- FH16: “acho que é mais pelo sotaque, né? Eu procurei analisar o sotaque e o assunto, da forma como ele falou e o assunto que ele estava falando. Eu penso que pessoas mais cultas falam de certos assuntos de uma forma diferente” [e como que elas falam?] “pelo tom da voz mesmo, né? Às vezes erro de português, talvez que eu percebi assim, puxando mais um r, um jeito mais caipira mesmo de falar.”
- AH18: “ah, não sei lembrar não. Tem algumas frases tipo assim: ‘eu fui lá’, entendeu? Coisa de gramática.”
- BM40: “teve uns que não falou o s... venda de veículo, ‘trabalho em loja de... coisa de venda de veículos’, aí falou ‘venda de veículo’. Em outros eu achei a entonação da voz assim meio caipira...” [tipo o quê, você lembra?] “ah... não”
- AH34: “falta de interesse, algumas que eu falei ‘pouco’ foi por falta de interesse da pessoa”
- DM16: “a palavra que ele falava” [tipo o quê?] “começava a falar certo, assim... depois ele mudava, sabe?” [mudava o quê?] “ah, assim, ah... num sei te falar não. Ele falava ‘belorizonte’ errado.”
- DM35: “às vezes é a fala, é... assim, engolindo letras, é... um que falou bastante ‘ídolo’, né?” [a pessoa falava o quê?] “falava ‘aquela pessoa daquele filme é meu ídolo’, acho que é isso, não?”
- IM15: “o jeito que a pessoa falou assim... e as palavras, às vezes, tipo... engolia algumas letras” [tipo o quê?] “‘aqueḷ’, é... ‘aqueḷ menino’ ”
- EM39: “você corrigiu algumas palavras... numas você falou caipira pra mostrar que não usa o vocabulário direito, no outro você já corrigiu o vocabulário. Às vezes a frase é a mesma, mas a maneira de falar é diferente.”
- LM40: “porque alguns sempre têm alguma coisa mal pronunciada ou errado” [mas o quê? Você lembra de alguma coisa?] “‘sécu passado’... é século. É... ‘eu estudo numa escola aqui pertinho de casa’, num precisa disso não uai, ‘eu estudo numa escola aqui perto’.”

- LM16: “por exemplo, naquele que fala ‘veículo’, sabe? Ah... tinha uns assim, sabe? ‘tá’, eu fiquei meio assim em dúvida, coloquei mesmo o que eu achava.”
- NM16: “eu olhei muito foi o sujeito da frase, sabe?” [como assim?] “o que deu pra analisar na frase, tipo análise sintática e morfológica da frase, eu fiz foi assim. Teve uma que foi assim ‘eu gosto das duas, mas eu gosto mais daquela’, eu acho que esse ‘mas eu gosto mais’, isso tá errado.”
- TH18: “tem uns que repete, só que tem muitas palavras por exemplo que a tonalidade..., por exemplo, ‘com eɣ’, eu olhei muito isso. ‘século passado’, por exemplo. ‘música do sécu passado’. É... num deu a tonalidade boa também não”
- CH33: “parece que a pessoa às vezes deixa de pronunciar o s, ou, como é que fala? Tem um jeito de falar assim do interior, meio caipira, tipo ‘pioɹ’, uma coisinha assim”

As respostas acima indicam que grande parte dos informantes não soube dizer porque atribuiu baixa ou alta escolaridade ao falante pela forma como a frase foi dita. Alguns identificaram a velarização da lateral, o apagamento da sílaba final, ou um falar ‘caipira’, como características da fala de uma pessoa que estudou pouco. De uma maneira geral, parece-nos que os falantes têm pouca consciência dos fenômenos estudados neste trabalho. Os resultados do teste de avaliação das variantes podem trazer alguns indícios sobre o grau de consciência de tais fenômenos.

Para a análise do teste de avaliação, foi criado um banco de dados com uma variável dependente composta pelas 4 variantes analisadas neste estudo e uma variável independente, codificada como:

Variável *avaliação*

0. estudou muito

1. estudou pouco

A partir desse procedimento, foram codificadas 42 frases para cada um dos 16 informantes, o que gerou um banco de dados com 672 ocorrências. A partir desse conjunto de dados, poderíamos avaliar a associação que os informantes faziam entre a produção de uma determinada variante e a escolaridade de quem a produziu.

A tabela abaixo apresenta o resultado das respostas dadas pelos informantes para cada uma das frases avaliadas:

Tabela 61: Frases distribuídas de acordo com a avaliação dos falantes

Avaliação Frases	Muito		Pouco		Total N
	n	%	n	%	
A Daniela falou que o Marcelo tem cara-de-caval (cavalo)	7	43,75	9	56,25	16
A Daniela falou que o Marcelo tem cara-de-cavalu. (cavalo)	9	56,25	7	43,75	16
A mulher do Seu João adora ficar na janeL. (janela)	12	75,00	4	25,00	16
A mulher do Seu João adora ficar na janeLθ. (janela)	14	87,50	2	12,50	16
A paneL que tá lá no fogão tá queimando. (panela)	8	50,00	8	50,00	16
A paneLθ que tá lá no fogão tá queimando. (panela)	11	68,75	5	31,25	16
Aque menina comprou um bicho. (aquela)	4	25,00	12	75,00	16
AqueLθ menina comprou um bicho (aquela)	9	56,25	7	43,75	16
Aquel menino correu atrás da passarinho. (aquele)	3	18,75	13	81,25	16
Aquel i menino correu atrás da passarinho. (aquele)	8	50,00	8	50,00	16
Belo Horizonte fica num vaĴ. (vale)	6	37,50	10	62,50	16
Belo Horizonte fica num vaLi. (vale)	3	18,75	13	81,25	16
Ê falou que não te conhece não. (ele)	5	31,25	11	68,75	16
El i falou que não te conhece não. (ele)	8	50,00	8	50,00	16
É saiu cedo e foi lá em casa sozinha. (ela)	3	18,75	13	81,25	16
Elθ saiu cedo e foi lá em casa sozinha. (ela)	5	31,25	11	68,75	16
Eu estudo numa escó pertinho daqui de casa. (escola)	2	12,50	14	87,50	16
Eu estudo numa escolθ pertinho daqui de casa. (escola)	7	43,75	9	56,25	16
Eu fui lá na casa do beto pra conversar com eĴ. (ele)	6	37,50	10	62,50	16
Eu fui lá na casa do beto pra conversar com eli. (ele)	5	31,25	11	68,75	16
Eu gostei dos dois, mas eu gostei mais daqueĴ (daquele)	9	56,25	7	43,75	16
Eu gostei dos dois, mas eu gostei mais daquel. (daquele)	12	75,00	4	25,00	16
Eu só gosto de andar de veícu novo. (veículo)	6	37,50	10	62,50	16
Eu só gosto de andar de veículu novo. (veículo)	12	75,00	4	25,00	16
Eu trabalho é com venda de veícuL (veículo)	6	37,50	10	62,50	16
Eu trabalho é com venda de veícuĴ (veículo).	5	31,25	11	68,75	16
Maria toma pílu pra num engravidar. (pílula)	12	75,00	4	25,00	16
Maria toma píluLθ pra num engravidar. (pílula)	10	62,50	6	37,50	16
Nossa... a peĴ deĴ tá toda queimada de sol. (pele dele)	11	68,75	5	31,25	16
Nossa... a peli deli tá toda queimada de sol. (pele deli)	13	81,25	3	18,75	16
O ator principal desse filme é meu ido. (ídolo)	3	18,75	13	81,25	16
O ator principal desse filme é meu ídol u. (ídolo)	4	25,00	12	75,00	16
O menino pulou igualzinho um gril u. (grilo)	14	87,50	2	12,50	16
O menino pulou igual um gril. (grilo)	8	50,00	8	50,00	16
O cantor do restaurante só canta música do sécu passado. (século)	4	25,00	12	75,00	16
O cantor do restaurante só canta música do sécul u passado.	5	31,25	11	68,75	16
O menino foi pra escoĴ (escola)	7	43,75	9	56,25	16
O menino foi pra escolθ. (escola)	11	68,75	5	31,25	16
Quando o menino viu a cara dela ele ficou trêmu. (trêmulo)	6	37,50	10	62,50	16
Quando o menino viu a cara dela ele ficou trêmul u. (trêmulo)	12	75,00	4	25,00	16
Vamos almoçar naquele restaurante de comida a quil? (quilo)	11	68,75	5	31,25	16
Vamos almoçar naquele restaurante de comida a quil u? (quilo)	4	25,00	12	75,00	16
Total	320	47,62	352	52,38	672

Para tentar identificar se os falantes tinham ou não consciência social dos fenômenos analisados em Itaúna, optei inicialmente por agrupar as variantes [l], [ɫ] e \emptyset em oposição à variante [l]_{+vogal}. O resultado da regressão logística com seleção por *stepwise* pode ser visto na tabela abaixo:

Tabela 62: Resultados do efeito da variável *avaliação* na variável dependente
[l]_{+vogal} ~ [l]/[ɫ]/ \emptyset

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Avaliação	Muito*	151 / 320	47,2	0,010	1,0
	Pouco	201 / 352	57,1		1,5

* fator de referência

O resultado da tabela acima indica que, se a frase produzida continha uma variante diferente de [l]_{+vogal}, a chance do falante associar tal ocorrência a alguém que estudou pouco é 1,5 vezes a chance de associá-la a alguém que estudou muito. Tal resultado apresenta indícios de que, no teste aplicado, os falantes associaram a produção de variantes diferentes da forma plena a pessoas de baixa escolaridade.

O passo seguinte foi realizar o mesmo procedimento, subdividindo o banco de dados entre os processos analisados neste trabalho. A primeira variável dependente analisada foi o apagamento da sílaba formada por /l/_{+vogal}.

Tabela 63: Resultados do efeito da variável *avaliação* na variável dependente
[l]_{+vogal} ~ \emptyset

	Fatores	n ₁ / n _t	% ₁	p-valor (Wald)	Razão de chances
Avaliação	Muito*	45 / 214	21,0	<0,001	1,0
	Pouco	99 / 250	39,6		2,5

* fator de referência

O resultado da tabela acima indica que a chance do falante associar a produção da variante \emptyset a alguém que estudou pouco é 2,5 vezes a chance de associá-la a alguém que estudou muito. Assim, parece termos indícios de que os falantes associaram tal variante a pessoas de baixa escolaridade, indício de um certo estigma da variante.

A tabela abaixo apresenta o resultado para a variável dependente [l]_{+vogal} ~ [l]. Para a seleção da variável *avaliação* utilizei o método *enter*, já que a variável *avaliação* não foi selecionada pelo método *forward*.

Tabela 64: Resultados do efeito da variável *avaliação* na variável dependente [1]+vogal ~ [1]

	Fatores	n_1 / n_t	% ₀₁	p-valor (Wald)	Razão de chances
Avaliação	Muito*	67 / 236	28,4	0,928	1,0
	Pouco	61 / 212	28,8		1,0

* fator de referência

O resultado apresentado na tabela acima indica que não houve associação entre a produção da variante [1] e a escolaridade de quem a produziu. Esse resultado pode ser um indício de que o fenômeno de apagamento da vogal está abaixo do nível de consciência social do fenômeno na cidade de Itaúna.

Resultado semelhante pode ser observado na análise da variável dependente [1]+vogal ~ [ɣ]. A tabela abaixo apresenta os resultados de um modelo de regressão com seleção da variável independente *avaliação* feita pelo método *enter*.

Tabela 65: Resultados do efeito da variável *avaliação* na variável dependente [1]+vogal ~ [ɣ]

	Fatores	n_1 / n_t	% ₀₁	p-valor (Wald)	Razão de chances
Avaliação	Muito*	39 / 106	36,8	0,614	1,0
	Pouco	41 / 102	40,2		1,2

* fator de referência

O resultado apresentado acima indica que não houve associação entre a produção da variante velarizada e a escolaridade de quem a produziu. Semelhante ao processo de apagamento da vogal, tal resultado pode ser um indício de que o fenômeno de velarização está abaixo do nível de consciência social do fenômeno na cidade de Itaúna e não indicou estigma considerável.

CAPÍTULO 7

CONCLUSÕES FINAIS

Este trabalho apresentou a análise do fenômeno variável envolvendo itens lexicais com sílaba final átona formada por /l/vogal na cidade de Itaúna/MG. O objetivo principal foi identificar e analisar a influência de determinadas variáveis independentes nas variantes encontradas na fala da comunidade estudada. Os pressupostos da *sociolinguística variacionista* formaram parte do arcabouço teórico-metodológico utilizado na condução desta pesquisa. Além disso, foram considerados os pressupostos da *fonologia de uso*, com atenção especial à interferência da frequência de ocorrência do item lexical nos processos de variação e mudança lingüística.

A identificação das variantes em análise pôde ser feita a partir da análise acústica dos dados de fala coletados em Itaúna. A análise acústica, dentre outras características, trouxe-nos indícios de que os processos variáveis analisados nesta pesquisa podem ser caracterizados como foneticamente graduais, dada a dificuldade de categorização em algumas ocorrências das variantes consideradas e a presença de variantes intermediárias, cuja produção encontrava-se entre duas variantes categorizadas. Tais variantes ocorreram, por exemplo, no caso do aquéa, para aquela, ou nos casos em que o sinal acústico apresentava ausência de segmentos que pareciam estar presentes considerando-se critérios perceptivos, ou ainda, nos casos em que o segmento vocálico seguinte à lateral apresentava duração muito reduzida em relação a outras vogais na mesma posição. Na seção 2.2 foram apresentados os processos variáveis analisados neste estudo, sendo: 1. apagamento da sílaba final átona /l/+vogal; 2. apagamento da vogal da sílaba final átona /l/+vogal e 3. apagamento da vogal da sílaba final átona /l/+vogal e velarização da lateral.

No capítulo 4, apresentou-se o modelo estatístico de regressão logística, utilizado nos estudos variacionistas e implementado no software *Varbrul* a partir da versão desenvolvida no final da década de 1970. Neste capítulo, foi possível mostrar que a utilização de um método convencional de codificação dos fatores, chamado *fator de referência*, diferente do método implementado no Varbrul, denominado *desvio da média*, não altera significativamente os resultados finais. A utilização do método de estimação dos efeitos dos fatores baseado na ponderação da ocorrência relativa do fator no Varbrul não

apresentou resultados significativamente diferentes para nenhum dos fatores considerados nesta pesquisa em relação ao método convencional de estimação dos efeitos. Isso indica que, pelo menos neste estudo, a utilização do Varbrul ou de qualquer outro software que gere resultados de *regressão logística*, como por exemplo, o SPSS ou o R, gera resultados estatísticos semelhantes. A utilização de um método estatístico não-convencional motivada pela presença de dados mal distribuídos entre os fatores no banco de dados, conforme Sankoff (1988), também não apresentou resultados diferentes, no banco de dados analisado, em relação ao método convencional. A opção pelo SPSS se deu, entre outras coisas, pela possibilidade de identificação da significância entre os fatores em uma variável, possibilitada pelo *teste de Wald*; pela possibilidade de utilização de outros métodos, como o CART, ou o *coeficiente de correlação de Spearman*; pela facilidade de identificação da *interação* entre variáveis independentes e pela possibilidade de incluir variáveis independentes com muitos fatores, como no caso da variável *item lexical*, que apresentava 150 fatores.

Os resultados apresentados nos testes de interação, seção 6.2, mostraram a importância de se incluir o *termo da interação* entre variáveis gênero e faixa etária no modelo final, de forma que os resultados sejam mais fiéis à ocorrência dos dados observados. A identificação das interações entre gênero e faixa etária sugere que tais variáveis sejam analisadas simultaneamente, já que o efeito de uma na variável dependente, influencia o efeito da outra. Neste trabalho, a análise simultânea foi feita com o agrupamento das variáveis gênero e faixa etária.

A contabilização final das ocorrências das variantes mostrou que a variante \emptyset apresenta o maior percentual de ocorrência na fala dos informantes entrevistados (35,6%), seguida das variantes [1]_{+vogal} (31,6%), [1] (29,3%) e [ɛ] (3,6%).

A análise das variáveis gênero e faixa etária indicou que o fato do indivíduo ser jovem e do gênero masculino favorece a utilização das variantes mais inovadoras, para todos os processos variáveis analisados. O processo de velarização, especificamente, parece tratar-se de um marcador social do grupo formado por jovens do gênero masculino. Os resultados indicaram que apesar do grupo social ser determinante, o comportamento individual também determina o uso das formas em variação na cidade de Itaúna; isso foi mencionado também em Oliveira (1991, 1992, 1995 e 1997), em estudos sobre o dialeto falado em Belo Horizonte. Como pôde ser visto, uma amostra bastante homogênea, composta de 16 indivíduos, com 4 indivíduos em cada célula, apresentou resultados para os reagrupamentos dos indivíduos, se considerarmos o efeito individual na variação, que não

refletem os resultados obtidos a partir de uma estratificação social analisada neste trabalho. Tal resultado pode indicar que características sociais ou individuais não mensuradas neste estudo podem trazer respostas aos questionamentos levantados com relação ao comportamento lingüístico dos indivíduos.

A análise das variáveis independentes lingüísticas apresentou resultados bastante relevantes para a análise da variabilidade ocorrida nos itens analisados. O processo de apagamento da vogal é favorecido por um contexto vocálico seguinte. Tal processo pôde ser caracterizado como um processo mais geral, já que atua em diferentes classes. O processo de *apagamento da sílaba* trata-se de um processo mais restrito aos itens pronominais, mas parece também estar se alastrando para classes não-pronominais, como pôde ser visto por exemplo nas ocorrências de *escó*, *para escola*, e *fá*, *para fala*, entre outros. O favorecimento do contexto seguinte *consoante*, quando todos os itens são incluídos na análise, não se confirma quando são analisados somente itens não-pronominais. Uma explicação para o favorecimento do contexto seguinte *consoante* no processo de *apagamento da sílaba* exigiria um estudo mais aprofundado, em que fosse analisada cada uma das ocorrências individualmente, para que se pudesse, a partir daí, identificar a causa deste favorecimento. O favorecimento do contexto seguinte *pausa* no processo de velarização da lateral mostra que esse fenômeno se destingue dos processos de apagamento da sílaba e de apagamento da vogal, pois, na velarização, não há um processo de redução causada pela coarticulação. O favorecimento causado pela *pausa* pode ser mais um indício de que a variante velarizada está relacionada a uma marcação social do grupo masculino-jovem, já que sua presença numa posição final de frase coloca tal variante em proeminência.

Com relação à variável *vogal da variável*, foi possível mostrar que a altura da vogal influencia no seu apagamento, já que vogais mais altas são mais apagadas.

Na seção 6.6.3, foram apresentados indícios da interferência da freqüência de ocorrência do item na variação ocorrida na fala de Itaúna. Os testes realizados indicaram que a chance de um item de alta freqüência de ocorrência sofrer algum processo de redução corresponde a 3 vezes a chance de um item de baixa freqüência sofrer tal processo. Esse resultado corrobora a hipótese da *fonologia de uso* de que, em processos foneticamente motivados, os itens de alta freqüência de ocorrência estão mais sujeitos a processos de mudança do que os itens de baixa freqüência. Os resultados que indicam a classe de pronomes como favorecedora do apagamento da sílaba podem sugerir também que a alta freqüência desses itens na fala dos informantes, como pode ser comprovado no anexo 5,

seja uma das responsáveis pelo processo de apagamento da sílaba. Há uma alta frequência de ocorrência dos pronomes pessoais *ele/eles/ela/elas* seguidos de verbo, o que poderia explicar um maior uso da variante \emptyset neste contexto. Entretanto, somente isso não explicaria o maior cancelamento da sílaba quando seguido por verbos auxiliares. Como pôde ser visto na seção 6.5.5, agrupamentos compostos por pronome+auxiliar, nos dados coletados, não são mais frequentes do que estruturas do tipo pronome+verbo. É preciso aprofundar os estudos neste sentido. Além da interferência da frequência, existe ainda a possibilidade de atuação de um fator sintático ou semântico envolvendo o processo de apagamento da sílaba.

A presença de /S/ seguinte na sílaba formada por /l/+vogal também apresentou efeitos estatisticamente significativos para a análise dos processos variáveis. Os resultados indicam que a presença do /S/ favorece o cancelamento da sílaba formada por /l/+vogal e desfavorece o cancelamento da vogal final. Com relação ao apagamento da sílaba, a análise da interferência da variável independente *indeterminação* na análise dos dados do pronome *eles* mostrou que o fato de tal pronome ser utilizado como indeterminador do sujeito não favorece significativamente o apagamento da sílaba, no conjunto de dados analisado. Já que outros itens também mostraram maior apagamento da sílaba quando seguida de /S/, neste trabalho, atribui-se tal resultado (1) a uma quantidade maior de material fonético na sílaba seguida de /S/, o que permite que o sentido do item seja mais facilmente retomado ou (2) a um apagamento da sílaba medial em proparoxítonas, considerando o *princípio de licenciamento da coda*. O desfavorecimento da variante [1] na presença do /S/ seguinte no fenômeno de apagamento da vogal pode ser atribuído à presença de um padrão fonotático, estabelecido probabilisticamente, conforme Bybee (2001, p. 13), que tende a desfavorecer a seqüência [1s] na língua portuguesa.

Com relação à variável *tonicidade*, há um favorecimento de variantes mais reduzidas quando o item é proparoxítono. Tal fato já havia sido observado em Martins (2001). Entretanto, outros estudos precisariam ser feitos nesse sentido, já que os resultados observados para o apagamento da sílaba podem estar sendo muito influenciados pelo item *Divinópolis*.

A análise da variável *estilo* mostrou que o fenômeno de apagamento da vogal é um fenômeno mais geral e mais abrangente do que os fenômenos de apagamento da sílaba e de velarização.

Os resultados apresentados no *teste de avaliação* indicam que nos fenômenos de apagamento da vogal e de velarização não há associação entre a produção das variantes [ɫ] e [ɮ] e a escolaridade de quem as produziu. Isso pode indicar que tais fenômenos não sejam socialmente estigmatizados e que, além disso, podem estar abaixo do nível de consciência social na cidade de Itaúna. Com relação ao *apagamento da sílaba*, ao contrário das variantes [ɫ] e [ɮ], pode ser identificada uma associação, feita pelos entrevistados, entre a variante Ø e a baixa escolaridade do falante. O que revela estigma social para tal variante e a possibilidade desse fenômeno estar acima do nível de consciência social.

A análise dos itens lexicais indicou a necessidade e a importância de se fazer um estudo mais amplo, item a item, analisando conjuntamente os efeitos gerados pelas variáveis favorecedoras, a frequência e o indivíduo.

Este trabalho apresentou contribuições para o estudo dos processos de variação e mudança sonora no português do Brasil, especialmente relacionados ao processo de variação ocorrido nos itens terminados em sílaba átona formada por /l/+vogal. Puderam ser apresentadas evidências de que características associadas ao uso, como a frequência de *token*, parecem também atuar, assim como fatores internos, nos fenômenos de etiologia fonética, caracterizados como “*regular sound change*” nos termos de Labov (1994); evidenciando-se assim uma seletividade lexical. Os fenômenos variáveis analisados neste estudo puderam ser caracterizados de acordo com suas especificidades. O fenômeno de velarização da lateral apresenta características bem próprias e há indícios de que ele esteja associado à marcação social do grupo formado por jovens do gênero masculino. O fenômeno de apagamento da vogal pôde ser caracterizado como um fenômeno de etiologia fonética, mais geral e mais abrangente do que o apagamento da sílaba e a velarização. No fenômeno de apagamento da sílaba, além das variáveis consideradas neste estudo, pode ser que outros fatores não controlados aqui estejam atuando. É possível que o apagamento da sílaba seja gradual e que haja o cancelamento do /l/ e o apagamento da vogal, não necessariamente nesta ordem. Um estudo mais aprofundado do apagamento da sílaba pode trazer algumas contribuições interessantes para entendermos mais claramente os fatores que atuam sobre ele.

Acredito que trabalhos realizados em outros dialetos possam oferecer contribuições importantes para uma melhor compreensão deste processo variável no PB. Acredito também que estudos que apresentem um maior aprofundamento dos modelos estatísticos

sejam relevantes, a fim de que se possa avaliar com mais propriedade o uso de modelos tendo em vista a especificidade dos dados a serem analisados. Outras propostas interessantes, a meu ver, relacionam-se a um aprofundamento na identificação e análise das características acústico-articulatórias do segmento lateral especificamente, como realizado em Andrade (1999) e Sproat e Fujimura (1993). Tais estudos poderiam trazer indícios que contribuam na caracterização dos fenômenos que incidem sobre tal segmento. Acredito que estudos que mensurem características sociais e individuais mais específicas, além das características comumente analisadas nos estudos variacionistas, podem contribuir para que se identifique com mais clareza o conjunto de características individuais que determinam a maneira como um indivíduo fala e os fatores que determinam a propagação das mudanças lingüísticas.

REFERÊNCIAS

ABAURRE, Maria Bernadete M. *Aspectos da organização rítmica em português europeu e brasileiro: uma análise do acento secundário baseada em hierarquia de restrições*. Colóquio do Projeto Temático Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística. Campinas, 2002.

ALBANO, E. C. *Segment Frequency and Word Structure in Brazilian Portuguese*. XIII International Congress of Phonetic Sciences, 3, p. 346-349, 1995.

ALBANO, E. C. Criatividade e gradiência num léxico sem derivações. In: L. GRIMM-CABRAL e MORAIS. (Org.). *Investigando a linguagem: ensaios em homenagem a Leonor Scliar-Cabral*. Florianópolis-SC, p. 35-54, 1999.

ANDRADE, A. *Variação fonética de /l/ em ataque silábico em Português europeu*. XIII Encontro da APL, Lisboa, p. 55-76, 1997.

ANDRADE, Amália. On /l/ velarization in European Portuguese. In OHALA, J; HASEGAWA, Y; OHALA, M; GRANVILLE, D. & BAILEY, A. (ed.). *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, p. 543-546, 1999.

BAYLEY, R.. The quantitative paradigm. In CHAMBERS, J. K.; TRUDGILL, P. and SCHILLING-ESTES N. (eds), *The Handbook of Language Variation and Change*. Oxford: Blackwell, p. 117-141, 2002.

BIDERMAN, Maria Tereza Camargo. Frequency Dictionary of Contemporary Portuguese. In *Journal des Traducteurs*, 41, 2, p. 275-278, 1996.

BISOL, L. (org.). *Introdução a Estudos de Fonologia do Português Brasileiro*. Edupcrs: Porto Alegre, 2001.

BYBEE, Joan L; HOPPER, Paul. *Frequency and the emergence of linguistic structure*. Amsterdam; Philadelphia, PA: John Benjamin's Pub. Co., 2001.

BYBEE, Joan. Lexical diffusion in regular sound change. In RESTLE and ZAEFFERER (eds.) *Sounds and Systems: Studies in Structure and Change*. Berlin: Mouton de Gruyter, p. 58-74, 2003..

BYBEE, Joan. *Phonology and language use*. Cambridge: Cambridge University Press, 2001.

BYBEE, Joan. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. In *Language Variation and Change* 14. p. 261-290, 2002.

CAMARA JR, J. Mattoso. *Historia e estrutura da língua portuguesa*. 4.ed. Rio de Janeiro: 1985.

CEDERGREN, H. and SANKOFF, D.. Variable rules: Performance as a statistical reflection of competence. *Language* 50, p. 233-255, 1974.

CHAMBERS, J.K. *Sociolinguistic Theory*. Oxford/Cambridge, Blackwell, 1995.

CONSELHO NACIONAL DE SAÚDE. Diretrizes e Normas Regulamentadoras de Pesquisas Envolvendo Seres Humanos: Resolução 196/96. Acesso em 02 de junho de 2005. Disponível em <<http://www.bioetica.ufrgs.br/res19696.htm>>.

CORRÊA, L. T. *A Forma Clítica de Pronome Pessoal no Dialeto Mineiro: uma variante sociolinguística*. Belo Horizonte: Dissertação de mestrado. Faculdade de Letras da UFMG, 1998.

COUTINHO, Ismael de Lima. *Pontos de gramática histórica*. [18. ed.]. Rio de Janeiro: Ao Livro Técnico, 2004.

CRISTÓFARO-SILVA, T. Descartando fonemas: a representação mental na Fonologia de Uso In: HORA, Dermeval da e COLLISCHONN, Gisela (org). *Teoria Lingüística: Fonologia e outros temas*. Ed. Universitária.UFPB, p. 200-231, 2003.

CRISTÓFARO-SILVA, T. Difusão Lexical: Estudo de Casos do Português Brasileiro. In: Mendes, Eliana Amarante de M.; Oliveira, Paulo Motta e Benn-Ibler, Veronika (org). *O novo milênio: interfaces lingüísticas e literárias*. FALE: Belo Horizonte, p. 209-218, 2001.

CRISTÓFARO-SILVA, Thaïs. *Fonética e Fonologia do Português: roteiro de estudo e guia de exercícios*. São Paulo: Contexto, 1999.

CRISTÓFARO-SILVA, Thaïs; OLIVEIRA, Marco Antonio de. Lateral Vocalization in Brazilian Portuguese. In: *3rd United Kingdom Language Variation Conference*, York - UK. Abstract booklet UKLVC3. York - UK : Department of Language and Linguistic Science, 2001.

DOBSON, Annette J. *An introduction to generalized linear models*. London: 1990.

ESPIGA, Jorge. Alofonia de /L/ no sul do Rio Grande do Sul: aspectos fonéticos e fonológicos. In: MATEUS, Maria Helena & RODRIGUES, Celeste. *Teoria lingüística: fonologia e outros temas*. Ed. João Pessoa: UFPB/Editora Universitária, p. 251-269, 2003.

FERREIRA, Carlos A., SOARES, José F., e CRUZ, Frederico R. B. Reconhecimento de Padrões em Estatística: Uma Abordagem Comparativa. *Proceedings of the V Brazilian Conference on Neural Networks - V Congresso Brasileiro de Redes Neurais*, Rio de Janeiro. p. 409–414, 2001

FISCHER, Graciana Simoní. *Um ambiente virtual multimídia de ensino na WEB, com transmissão ao vivo e interatividade*. Dissertação Ciência da Computador, UFRS, 2000.

FISHER, John L. Influências sociais na escolha de variantes lingüísticas. Trad. Elba I. Souto. In: FONSECA, Maria Stella, NEVES, Moema F. (org.). *Sociolingüística*. Rio de Janeiro: Eldorado, p. 87-98, 1974.

FUNDAÇÃO JOÃO PINHEIRO. Disponível em: <http://www.fjp.gov.br/>. Acesso em 07 de novembro de 2005.

GICK, Brian. Articulatory Correlates of Ambisyllabicity in English Glides and Liquids. In *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*, ed. J. Local, R. Ogden, and R. Temple. Cambridge: Cambridge Univ. Press, 1998.

GUY, G. R. Advanced VARBRUL analysis. In FERRARA, K., BROWN, B., WALTERS, K., and BAUGH J. (eds). *Linguistic Contact and Change*. Austin: University of Texas Department of Linguistics, p. 124-136, 1988.

GUY, G. R. Variation in the group and the individual: The case of final stop deletion. In LABOV, W. (ed). *Locating Language in Time and Space*. p. 1-36. New York: Academic Press, 1980.

GUY, Gregory e ZILLES, Ana M. S. *Análise Quantitativa em Sociolinguística*. [Manuscrito não publicado].

HOSMER, David W.; LEMESHOW, Stanley. *Applied logistic regression*. 2nd ed. New York: Wiley, 2000.

HUFFMAN, Marie K. *Phonetic variation in intervocalic onset /l/'s in English*. Journal of Phonetics, 25/2, p. 115-141, 1997.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Disponível em: <<http://www.ibge.gov.com>>. Acesso em: 07 de novembro de 2005.

JOHNSON, Keith. *Quantitative Methods in Linguistics*. Disponível em <<http://linguistics.berkeley.edu/~kjohnson/quantitative/>>, 2004.

KAYE, Jonathan. "Coda" licensing. *Phonology Yearbook*: o. 301-330, 1989.

KENT, R. D. and READ, C. *The Acoustic of Speech. Singular*. San Diego, 1992.

KIPARSKY, P. The phonological basis of sound change. In GOLDSMITH, J. A. (ed.) *The handbook of phonological theory*. Oxford: Blackwell, p. 640-670, 1995.

KLEINBAUM, David G. *Logistic regression: a self-learning text*. New York: 1994.

LABOV W. Stages in the acquisition of standard English. In SHUY, R. (ed.). *Social Dialects and Language Learning*. Champaign, Ill.: National Council of Teachers of English, 1964.

LABOV, W. *Driving forces in linguistic change*. 2002 Disponível em: <<http://www.ling.upenn.edu/~wlabov/Papers/DFLC.htm>>.

LABOV, W. *Principles of Linguistic Change: internal factors*. Oxford: Black Well, 1994.

LABOV, W. *Principles of Linguistic Change: social factors*. Oxford: Black Well, 2001.

LABOV, William. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LADEFOGED, Peter and MADDIESON, Ian. *The sounds of the World's Languages*. Oxford: Blackwell, 1996.

LAEL. Banco de dados do português. São Paulo: PUC/SP. Disponível em: <www.lael.pucsp.br/corpora>. Acesso em 10 maio de 2005.

LEECH, Nancy L., BARRETT, Karen C. and MORGAN, George A. *SPSS for intermediate statistics: use and interpretation*. Second Edition. New Jersey: Lawrence Erlbaum Associates, Publishers, 2005.

LEECH, Nancy L., BARRETT, Karen C. and MORGAN, George A. *SPSS for introductory statistics: use and interpretation*. Second Edition. New Jersey: Lawrence Erlbaum Associates, Publishers, 2005.

MAGALHÃES, José Olímpio de. Possibilidades fonéticas/fonológicas para as variações de 'ele, ela, eles, elas'. In: *Reunião Preparatória da XX ENANPOLL*. Ciclo de Palestras em Sociolinguística. Belo Horizonte, 2005.

MAIA, Clarinda de Azevedo. *História do Galego-Português*. Estado Lingüístico da Galiza e do Noroeste de Portugal desde o Século XIII ao Século XVI. Coimbra, INIC, 1986.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Técnicas de pesquisa*. 3. ed. rev. e ampl. São Paulo: Atlas, 1996.

MARTINS, Raquel Márcia Fontes. *O cancelamento da líquidas /l/ e /r/ intervocálicas no português contemporânea de Belo Horizonte*. Belo Horizonte: Dissertação de Mestrado. Faculdade de Letras da UFMG, 2001.

MATEUS, M. H. M. *Aspectos da Fonologia Portuguesa*. Lisboa: Centro de Estudos Filológicos 1975.

MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. 2nd. ed. London; New York: 1989.

MENDOZA-DENTON, N., HAY, J., and JANNEDY, S. Probabilistic sociolinguistics: beyond variable rules. in BOD, R., HAY, J. and JANNEDY, S. *Probabilistic Linguistics*. MIT Press, Cambridge, 2003.

MILROY, Lesley. *Language and social networks*. 2. ed. Oxford: B. Blackwell, 1987.

MORRISON, G. S. Dat is What the PM Said: A Quantitative Analysis of Prime Minister Chrétien's Pronunciation of English Voiced Dental Fricatives. *Cahiers linguistiques d'Ottawa*, 33. Ottawa, Ontario: University of Ottawa, Department of Linguistics, p. 1-21, 2005.

NOGUEIRA, Guaracy de Castro. *Itaúna em detalhes: enciclopédia ilustrada de pesquisa*. Itaúna: Jornal Folha do Povo, 2003.

OLIVEIRA, Marco Antônio de. Aspectos da difusão lexical. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 1, p. 31-41, 1992.

OLIVEIRA, Marco Antônio de. *Phonological variation and change in Brazilian Portuguese*. Tese (doutorado) - University of Pennsylvania, 1983.

OLIVEIRA, Marco Antônio de. Reanalizando O Processo de Cancelamento do (R) Em Final de Sílabas. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 6, n. 2, p. 70-97, 1997.

OLIVEIRA, Marco Antônio de. The Neogrammarian Controversy Revisited. *International Journal Of The Sociology Of Language*, Berlin, Alemanha, v. 89, n. 1, p. 93-105, 1991.

OLIVEIRA, Marco Antônio. O léxico como controlador de mudanças sonoras. *Revista de Estudos da Linguagem*, Belo Horizonte: Faculdade de Letras da UFMG, v. 36, p.75-92, 1995.

PAGANO, M. e GAUVREAU, K.. *Princípios de Bioestatística*. Ed. Thomson, 2ª Edição, SP, 2004.

PAIVA, Maria C. e DUARTE, Maria Eugênia L. *Mudança lingüística em tempo real*. Rio de Janeiro: Contra Capa Livraria, 2003.

PAOLILLO, John C. <http://ella.slis.indiana.edu/~paolillo>.

PHILLIPS, Betty S. Word frequency and the actuation of sound change. *Language*. 60, p. 320-342, 1984.

PICKETT, J. M. *The acoustic of speech communication*. Boston: Allyn and Bacon, 1999.

PINTZUK, S. *Varbrul programs [Computer program]*. Philadelphia: University of Pennsylvania Department of Linguistics, 1988.

RAND, D. and SANKOFF, D. *GoldVarb: A variable rule application for the Macintosh (version 3.0B)*. Montreal: Centre de recherches mathématiques, Université de Montréal, 1998.

ROUSSEAU P, SANKOFF D. A solution to the problem of grouping speakers. In: SANKOFF, D (ed.) *Linguistic Variation: Models and Methods*. Academic Press, New York, 1978.

ROUSSEAU, P. and SANKOFF, D. Advances in variable rule methodology. In SANKOFF, D. (ed.) *Linguistic variation: Models and methods*. New York: Academic, p. 57-69, 1978.

SAITO, H. Dependence and interaction in frequency data analysis in SLA research. *Studies in Second Language Acquisition*, 21, p. 453-475, 1999.

SANKOFF, D. Linguistics, Statistics In *Sociolinguistic Method And Linguistic Theory*, North-Holland Publishing Company And Pwn- p 74-82, 1982.

SANKOFF, D. Statistics in sociolinguistics. In MESTHRIE, R. (ed). *The Concise Encyclopedia of Sociolinguistics*, p. 828-834. London: Pergamon, 2001.

SANKOFF, D. Variable rules. In AMMON, U., DITTMAR, N., and MATTHEIER, K. J. (eds), *Sociolinguistics: An International Handbook of the Science of Language and Society*, vol. 2, p. 984-997. Berlin: Mouton de Gruyter, 1988.

SANKOFF, David. Statistics in linguistics. *New York Encyclopaedia of the statistical sciences 5*. Wiley. p.74-81, 1985.

SANKOFF, David; TAGLIAMONTE, Sali and SMITH Eric. *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto, 2005.

SARDINHA, T.B.. *The bank of Portuguese - A multi-million word monitor corpus of Brazilian Portuguese*. Trabalho apresentado no 4th Corpus Linguistics Symposium, InPLA 2001. PUC/SP, São Paulo, Brasil, 1991.

SARDINHA, Tony Berber. Corpus Linguistics: history and problematization. *DELTA*, vol.16, no.2, p.323-367, 2000.

SARDINHA, Tony Berber. *Linguística de Corpus*. Barueri, SP: Manole, 2004.

SAITO, Hidetoshi. Dependence and interaction in frequency data analysis in SLA research. *Studies in Second Language Acquisition* 21, p. 453-475, 1999.

SCOBIE, James M. and WRENCH, Alan A. An articulatory investigation of word final /l/ and /l/-sandhi in three dialects of English. In SOLÉ, M. J., RECASENS, D. & ROMERO J. (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*, Sydney, 2003.

SHEN, Z. Lexical diffusion: a population perspective and a mathematical model. *Journal of Chinese Linguistics*, [S.L], v. 18, p. 159-291, 1990.

SPROAT, R. & FUJIMURA O. Allophonic variation of English /l/ and its implication for phonetic implementation. *Journal of Phonetics*, 21: p. 291-311, 1993.

SPSS Inc. (2005). SPSS 13.0 [Computer software]. Chicago.

TASCA, Maria. Variação e Mudança do segmento lateral na coda silábica. In: BISOL, Leda e BRESCANCINI, Cláudia (Orgs.) *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, p. 269-302, 2002.

TESSYER, Paul. *História da Língua Portuguesa*. São Paulo: Martins Fontes, 1997.

VANDRESEN, Paulino. Variação e Mudança nas consoantes pós-vocálicas na região sul. In: CABRAL, L.G. e MORAIS, J. (Orgs.) *Investigando a linguagem*. Florianópolis: Editora Mulheres, 1999.

VIEGAS, Maria do Carmo. *O alçamento de vogais médias pretônicas e os itens lexicais*. Tese (doutorado) - Universidade Federal de Minas Gerais, 2001.

WILLIAMS, Edwin Bucher. *Do latim ao português: fonologia e morfologia históricas da língua portuguesa*. 3. ed. Rio de Janeiro: Tempo Brasileiro, 1975.

WRENCH, Alan A. and SCOBIE, James M.. Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In SOLÉ, M. J., RECASENS, D. & ROMERO J. (eds.) *Proceedings of the 6th International Seminar on Speech Production*, Sydney, 2003.

YOUNG, R., & BAYLEY, R. VARBRUL analysis for second language research. In BAYLEY, R. & PRESTON, D. R. (Eds.). *Second language acquisition and linguistic variation*, p. 253–306. Amsterdam: Benjamins, 1996.

YOUNG, R., & YANDELL, B. Top-down versus bottom-up analyses of interlanguage data: A reply to Saito. *Studies in Second Language Acquisition* 21, p. 477-488, 1999.

ZÁGARI, Mário R. L. *Fonologia diacrônica do Português*. Juiz de Fora: UFJF, 1988.

ZURMELY, Roland Marcel. *Digitalização de um sinal analógico*. Disponível em: <<http://paginas.terra.com.br/lazer/py4zbx/teoria/digitaliz.htm>>. Acesso em 25 de setembro de 2005.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)