

Raquel Márcia Fontes Martins

**A ORGANIZAÇÃO DO COMPONENTE FONOLÓGICO E O  
COMPORTAMENTO DO INDIVÍDUO**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Estudos Lingüísticos, da Faculdade de Letras da Universidade Federal de Minas Gerais (UFMG), como requisito parcial à obtenção do título de Doutora em Lingüística.

Área de concentração: Lingüística

Linha de pesquisa: D (Organização Sonora da  
Comunicação Humana)

Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Thaïs Cristófaró Silva

Belo Horizonte

Faculdade de Letras da UFMG

2007



Aos meus Pais, grandes responsáveis por  
esta e por todas as outras minhas  
conquistas!

## AGRADECIMENTOS

Queria ter *a palavra certa* para agradecer a todas as pessoas que participaram, de alguma forma, deste trabalho. Tentarei me expressar agradecendo principalmente:

À minha Orientadora, Profa. Dra. Thaïs Cristófaró Silva, quem muito admiro, respeito e a quem sou infinitamente grata pela competente orientação, por toda a dedicação e pelas minuciosas leituras deste texto. Thaïs, eu não saberia dimensionar todo o crescimento acadêmico, profissional e, inclusive, pessoal que tive convivendo com você nesses anos e porque não dizer, para ser mais exata, nessa última década! Você sempre foi e continua sendo um grande estímulo para mim! Obrigada por tudo!!!

Aos Professores César Reis, Maria do Carmo Viegas, Rui Rothe Neves e Thaïs Cristófaró Silva, pelas aulas nas suas disciplinas tão produtivas.

À Profa. Joan Bybee, por me permitir participar de seus cursos no período em que estive na Universidade do Novo México.

À Professora Adriana Marusso, pelas excelentes sugestões na elaboração do material de coleta de dados.

Ao Prof. Rui Rothe Neves, pelas contribuições no delineamento da metodologia deste estudo.

Às Professoras Christina Abreu Gomes e Heliana Mello, pelas contribuições dadas no Exame de Qualificação, as quais redirecionaram o meu trabalho de forma muito adequada.

Aos Professores Christina Abreu Gomes, Érica Parlato, Heliana Mello e Wilson Araújo de Carvalho, pelos preciosos comentários e sugestões dados na defesa desta tese.

Às amigas Ana Paula e Daniela... só vocês para me ouvirem falar tantas vezes do meu trabalho com tanta paciência e atenção! Ana Paula, de forma irônica nossa amizade se fortaleceu nos EUA e o importante é que, seja aqui ou em qualquer outra parte do mundo, temos uma grande afinidade que não há distância que desfaça. Daniela, apesar de termos nos conhecido antes, foi graças ao doutorado que nos tornamos grandes amigas. Você foi uma companheira maravilhosa todo esse tempo! Minha afinidade também com você transcende essa vida.

Ao Alan, amigo *expert* em estatística, pelas aulas, por me acolher quando estava desorientada na análise dos meus dados. Obrigada pela sua paciência e didática em me ensinar tudo com tanto cuidado.

Ao Leandro, obrigada também por sua ajuda com a estatística dos meus dados.

À Christine, obrigada pela amizade, carinho e apoio; por me animar a apresentar meu trabalho no Simpósio de Austin, quando eu não acreditava que isso seria possível. Valeu demais ter seguido o seu conselho!

Ao Leo Almeida, pela amizade, pelos momentos de descontração e por me socorrer em questões de informática. Leo, fez toda diferença a ajuda que você me deu com o *Layout* desta tese e também com a apresentação da defesa. Como aprendi! Obrigada mesmo!

Aos colegas de curso, principalmente, Adelma, Alan, Ana Paula, Camila, Daniela, Izabel, Kátia, Larissa, Máira, Rubens e Sandro, pelos momentos tão agradáveis que passamos nas disciplinas cursadas.

À minha família, quem mais me estimulou a sempre valorizar e buscar conhecimento. Mãe, Pai, Rudy, Isaac, Carol, Dedé, Mari, João Romeu, Luciano e Leo, obrigada por todo apoio que vocês me deram em todos esses anos de estudo!

Ao Wagner, pelo amor, apoio e compreensão, por entender a minha vontade e necessidade de estudar, aceitando os momentos em que estive ausente.

Aos participantes desta pesquisa, pela boa vontade e gentileza de participar do presente estudo. Minha mais sincera gratidão a vocês!

Ao CNPq, pela bolsa de estudos concedida durante o curso.

Individual differences “[...] reveal the range of possible state spaces the system can occupy [...]

Individual similarities, in turn, tell us about the constraints and limitations of that space.”

Thelen e Smith (1994, p. 145)

## LISTA DE FIGURAS

FIGURA 1 – Esquema com conexões lexicais (fonológicas) para [p], [pra] e [asa].	51
FIGURA 2 – Esquema com conexões lexicais (fonológicas e semânticas – relações morfológicas) para o morfema <i>-dor</i> .	52
FIGURA 3 – Nuvem de exemplares para o sufixo <i>-dor</i> .	59
FIGURA 4 – Nuvem de múltiplos exemplares para o sufixo <i>-dor</i> .	59
FIGURA 5 – Nuvem de exemplares para o sufixo <i>-dor</i> , com [-doh] e [-do] em competição (Baseado em CRISTÓFARO SILVA, 2002).	60
FIGURA 6 – Classificação em árvore (método CHAID) para as 30 palavras do estudo transversal.	139
FIGURA 7 – Categorias estilo informal e estilo formal em um contínuo de 0 (estilo informal) a 1 (estilo formal).	144
FIGURA 8 – Classificação em árvore (método CHAID) para a variável <i>Estilo</i> , em relação aos três fenômenos em análise ao mesmo tempo.	145
FIGURA 9 – Classificação em árvore (método CHAID) para a variável <i>Estilo</i> , em relação ao Fenômeno 1 (r #).	146
FIGURA 10 – Classificação em árvore (método CHAID) para a variável <i>Estilo</i> , em relação ao Fenômeno 2 (v-v).	147
FIGURA 11 – Classificação em árvore (método CHAID) para a variável <i>Estilo</i> , em relação ao Fenômeno 3 (cc).	147
FIGURA 12 – Classificação em árvore (método CHAID) para os indivíduos não-líderes somente, quanto aos três fenômenos em análise.	153
FIGURA 13 – Classificação em árvore (método CHAID) com as 30 palavras do estudo transversal somente para os indivíduos líderes.	154
FIGURA 14 – Classificação em árvore (método CHAID) com as 30 palavras do estudo transversal somente para os indivíduos não-líderes.	155
FIGURA 15 – Classificação em árvore (método CHAID) com a relação entre as variáveis <i>Liderança</i> e <i>Estilo</i> para os não-líderes.	159
FIGURA 16 – Classificação em árvore (método CHAID) com a relação entre as variáveis <i>Liderança</i> e <i>Estilo</i> para os líderes.	160
FIGURA 17 – Líderes e não-líderes quanto aos estilos informal (0) e formal (1).	162
FIGURA 18 – Classificação em árvore (método CHAID) para os 12 indivíduos (6 líderes e 6 não-líderes) quanto aos três fenômenos em análise.	171

FIGURA 19 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 1 (r #) apenas.....	172
FIGURA 20 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 2 (v-v) apenas.....	174
FIGURA 21 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 3 (cc) apenas.....	175
FIGURA 22 – Gradiência de fenômenos com relação ao Indivíduo 2 – L1V .....	180
FIGURA 23 – Indivíduo 1 – L1W e as 30 palavras em estudo.....	182
FIGURA 24 –Indivíduo 2 – L1V e as 30 palavras em estudo.....	182
FIGURA 25 – Indivíduo 3 – N1A e as 30 palavras em estudo .....	183
FIGURA 26 – Indivíduo 4 – N1L e as 30 palavras em estudo.....	183
FIGURA 27 –Indivíduo 5 – L2L e as 30 palavras em análise .....	184
FIGURA 28 – Indivíduo 6 – L2R e as 30 palavras em estudo.....	184
FIGURA 29 – Indivíduo 7 – N2L e as 30 palavras em estudo.....	185
FIGURA 30 – Indivíduo 8 – N2M e as 30 palavras em estudo.....	185
FIGURA 31 – Indivíduo 9 – L3R e as 30 palavras em estudo.....	186
FIGURA 32 – Indivíduo 10 – L3J e as 30 palavras em estudo.....	186
FIGURA 33 – Indivíduo 11 – N3Z e as 30 palavras em estudo.....	187
FIGURA 34 – Indivíduo 12 – N3G e as 30 palavras em análise.....	187
FIGURA 35 – Os 12 indivíduos no estilo fala espontânea .....	190
FIGURA 36 – Os 12 indivíduos no estilo tarefa de nomeação de figura.....	191
FIGURA 38 – Classificação em árvore (método CHAID) para as 16 palavras consideradas no estudo longitudinal.....	222
FIGURA 39 – Indivíduo – L2R e as 16 palavras consideradas no estudo longitudinal.....	225
FIGURA 40 – Indivíduo – N3Z e as 16 palavras consideradas no estudo longitudinal.....	225

## LISTA DE QUADROS

QUADRO 1: Os 3 fenômenos em estudo.....	23
QUADRO 2: Comparação entre a proposta tradicional e o Modelo de Exemplares .....	64
QUADRO 3: Informantes.....	75
QUADRO 4: Codificação referente ao QUADRO 3 .....	75
QUADRO 5: Líderes e não-líderes .....	77
QUADRO 6: Codificação dos participantes.....	79
QUADRO 7: As 50 palavras utilizadas no teste-piloto.....	86
QUADRO 8: As 30 palavras consideradas nesta pesquisa .....	108
QUADRO 9: Líderes e não-líderes quanto aos fenômenos em análise.....	176
QUADRO 10: Informações sobre corpus/ pesquisa e fenômeno para os 12 indivíduos .....	198

## LISTA DE TABELAS

TABELA 1: 15 palavras mais freqüentes: primeira seleção .....	90
TABELA 2: Dados quantitativos do Corpus NILC/ São Carlos .....	95
TABELA 3: Totais de palavras dos registros e tipos de textos do Corpus do LAEL ..	100
TABELA 4: Resultados da pesquisa de freqüência para as palavras “de” e “e” no LAEL-escrita e fala .....	101
TABELA 5: Freqüência das 15 palavras no Corpus analisado (corpora 1, 2 ou 3) e nos corpora do LAEL-escrita e fala e do NILC/ São Carlos.....	103
TABELA 6: As outras 15 palavras mais freqüentes do LAEL-escrita, distribuídas de acordo com os fenômenos dos corpora 1, 2 e 3.....	106
TABELA 7: Resultado geral dos dados do estudo transversal.....	130
TABELA 8: Análise da variável <i>Fenômeno</i> .....	132
TABELA 9: Análise da variável <i>Palavra</i> .....	135
TABELA 10: Análise da variável <i>Estilo</i> .....	142
TABELA 11: Análise da variável <i>Liderança</i> .....	149
TABELA 12: Relação das variáveis <i>Fenômeno</i> e <i>Liderança</i> : somente líderes .....	151
TABELA 13: Relação das variáveis <i>Fenômeno</i> e <i>Liderança</i> : somente não-líderes ....	152
TABELA 14: Análise da variável <i>Indivíduo</i> : dados percentuais .....	163
TABELA 15: Resultados de peso relativo para aos 12 indivíduos quanto aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em conjunto .....	165
TABELA 16: Classificação dos 12 indivíduos em faixas de peso relativo e quanto ao contínuo gradual não-líder (0) – líder (1) .....	168
TABELA 17: Resultados de peso relativo para os 12 indivíduos quanto aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em separado e em conjunto .....	177
TABELA 18: O comportamento dos 12 indivíduos no seu corpus de origem e no corpus atual .....	200
TABELA 19: Resultados do teste de proporção .....	201
TABELA 20: Dados do estudo longitudinal .....	205
TABELA 21: Percentual de cancelamento e peso relativo dos indivíduos do estudo longitudinal.....	208
TABELA 22: Indivíduos L2R e N3Z nos estudos transversal e longitudinal.....	212
TABELA 23: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Informante</i> : somente para o Indivíduo L2R .....	213

TABELA 24: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Informante</i> : somente para o Indivíduo N3Z .....	213
TABELA 25: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Coleta</i> : todas as 6 coletas, somente para o não-líder N3Z.....	215
TABELA 26: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Coleta</i> : todas as 6 coletas, somente para a líder L2R .....	215
TABELA 27: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Coleta</i> : sem a Coleta 6, somente para a líder L2R .....	216
TABELA 28: Tabela de contingência entre as variáveis <i>Cancelamento</i> e <i>Palavra</i> : as 16 palavras consideradas no estudo longitudinal .....	220

## SUMÁRIO

CAPÍTULO 1 .....	18
INTRODUÇÃO.....	18
CAPÍTULO 2 .....	26
O COMPORTAMENTO DO INDIVÍDUO.....	26
2.1 Introdução.....	26
2.2 O comportamento do indivíduo como unidade de análise .....	27
2.3 Irregularidades no comportamento do indivíduo e difusão gradual da mudança sonora na comunidade .....	33
2.4 Abordagens para fenômenos de mudança e variação sonora: comunidade lingüística, redes sociais e comportamento do indivíduo .....	36
2.5 Sumário.....	45
CAPÍTULO 3 .....	47
FUNDAMENTAÇÃO TEÓRICA .....	47
3.1 Introdução.....	47
3.2 Fonologia de Uso.....	47
3.3 Teoria de Exemplares .....	58
3.4 Lingüística Probabilística .....	65
3.5 Sumário.....	68
CAPÍTULO 4 .....	70
METODOLOGIA.....	70
4.1 Introdução.....	70
4.2 Tipo de método adotado .....	70
4.3 Amostra .....	74
4.4 Material.....	80
4.4.1 Discussão sobre os 3 corpora – 3 fenômenos do português brasileiro considerados .....	80
4.4.2 Entrevista e tarefas.....	82
4.4.3 Apresentação e análise das 30 palavras selecionadas neste estudo .....	83
4.4.3.1 O teste-piloto .....	84
4.4.3.2 Seleção inicial de 15 das 30 palavras em análise .....	89
4.4.3.2.1 Critérios de seleção das palavras da TAB. 1 .....	91
4.4.3.2.2 Avaliação, em corpora eletrônicos, dos dados obtidos dos corpora 1, 2 e 3 .....	94
4.4.3.2.2.1 Corpus CETEM/ Público - Corpus NILC/ Universidade de São Carlos..	95

4.4.3.2.2.2 Corpus LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala .....	99
4.4.3.2.2.3 Análise no Corpus NILC/ Universidade de São Carlos e no Corpus LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala .....	102
4.4.3.3 Apresentação dos outros 15 dados (obtidos do LAEL-escrita e fala) .....	105
4.4.3.3.1 Especificidades na seleção das palavras mais freqüentes do LAEL-escrita .....	107
4.4.3.4 Apresentação geral das palavras consideradas nesta pesquisa .....	107
4.5 Procedimentos de coleta e avaliação dos dados .....	109
4.5.1 Coleta do estudo transversal .....	110
4.5.1.1 Alguns esclarecimentos a respeito da coleta do estudo transversal.....	110
4.5.1.2 Dados .....	112
4.5.1.3 Conteúdo da entrevista de fala espontânea e das tarefas de nomeação por figura e de leitura.....	113
4.5.1.3.1 Entrevista de fala espontânea .....	113
4.5.1.3.2 Tarefa de nomeação por figura.....	113
4.5.1.3.3 Tarefa de leitura.....	114
4.5.1.3.4 Variáveis analisadas .....	119
4.5.2 Coletas do estudo longitudinal .....	122
4.5.2.1 Alguns esclarecimentos a respeito das coletas do estudo longitudinal .....	122
4.5.2.2 Dados .....	124
4.5.2.3 Variáveis analisadas .....	125
4.5.3 Equipamento de gravação utilizado.....	126
4.5.4 Esclarecimentos sobre a análise estatística realizada .....	126
4.6 Sumário.....	128
CAPÍTULO 5 .....	130
ESTUDO TRANSVERSAL: ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS .....	130
5.1 Introdução.....	130
5.2 Visão geral dos dados do estudo transversal .....	130
5.2.1 Análise da variável <i>Fenômeno</i> .....	132
5.2.2 Análise da variável <i>Palavra</i> .....	134
5.2.3 Análise da variável <i>Estilo</i> .....	142
5.2.4 Análise da variável <i>Liderança</i> .....	149
5.2.4.1 Relação entre as variáveis <i>Liderança</i> e <i>Fenômeno</i> .....	150

5.2.4.2 Relação entre as variáveis <i>Liderança</i> e <i>Palavra</i> .....	154
5.2.4.3 Relação entre as variáveis <i>Liderança</i> e <i>Estilo</i> .....	158
5.2.5 Análise da variável <i>Indivíduo</i> .....	163
5.2.5.1 Relação entre as variáveis <i>Indivíduo</i> e <i>Fenômeno</i> .....	170
5.2.5.2 Relação entre as variáveis <i>Indivíduo</i> e <i>Palavra</i> .....	181
5.2.5.3 Relação entre as variáveis <i>Indivíduo</i> e <i>Estilo</i> .....	189
5.3 Sumário.....	193
CAPÍTULO 6 .....	197
ESTUDO LONGITUDINAL: ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS .....	197
6.1. Introdução.....	197
6.2 Análise comparativa do comportamento dos 12 indivíduos no seu corpus de origem e na pesquisa atual.....	198
6.3 Análise e discussão dos resultados do banco de dados do estudo longitudinal..	205
6.3.1 Visão geral dos dados do estudo longitudinal .....	205
6.3.2 Análise das variáveis independentes <i>Indivíduo</i> , <i>Coleta</i> , <i>Palavra</i> .....	207
6.3.2.1 Análise da variável <i>Indivíduo</i> .....	207
6.3.2.1.1 Comparação dos indivíduos L2R e N3Z nos dois bancos de dados: transversal e longitudinal.....	211
6.3.2.2 Análise da variável <i>Coleta</i> .....	214
6.3.2.3 Análise da variável <i>Palavra</i> .....	217
6.3.2.3.1 Critérios para o recorte da variável <i>Palavra</i> .....	218
6.3.2.3.2 Análise das 16 palavras consideradas no estudo longitudinal.....	219
6.4 Sumário.....	226
CAPÍTULO 7 .....	229
CONSIDERAÇÕES FINAIS .....	229
BIBLIOGRAFIA .....	237
APÊNDICE A .....	251
Traduções nossas para as citações em língua inglesa.....	251
APÊNDICE B.....	254
Roteiro da entrevista de fala espontânea com as 30 palavras em estudo .....	254
APÊNDICE C.....	256
Tarefa de leitura.....	256
APÊNDICE D .....	258
Anotações a respeito das coletas e dos 2 participantes do estudo longitudinal.....	258

APÊNDICE E .....	260
Critérios de seleção das 50 palavras inicialmente consideradas neste trabalho .....	260
APÊNDICE F .....	269
Tabelas – palavras e indivíduos do estudo transversal .....	269
APÊNDICE G.....	275
Tabelas – palavras e indivíduos do estudo longitudinal.....	275

## RESUMO

Esta tese analisa o comportamento lingüístico de indivíduos específicos em relação a fenômenos fonológicos diferentes, avaliando a organização do componente fonológico. Sugere-se que a variabilidade individual atestada na produção da fala é incorporada nas representações lingüísticas. Adotam-se as seguintes teorias multirrepresentacionais: Fonologia de Uso (BYBEE, 2001), Teoria de Exemplos (PIERREHUMBERT, 2001b, 2003a) e Lingüística Probabilística (BOD; HAY; JANNEDY, 2003).

Nesta pesquisa, o conceito de líder da mudança lingüística proposto por Labov (2001, v. 2) é considerado. Dois tipos de estudo são realizados: um transversal com 12 indivíduos (6 líderes e 6 não-líderes) e um longitudinal de 6 coletas mensais com 2 desses 12 indivíduos (1 líder e 1 não-líder). Os resultados encontrados demonstram variações inter e intra-individuais relevantes e apontam para aspectos que estão em consonância com os modelos teóricos adotados neste trabalho. Dentre tais aspectos, podem-se citar: dinamicidade e plasticidade do léxico; padrões de difusão lexical da mudança sonora; caráter probabilístico da linguagem; léxico e gramática individuais no uso.

## ABSTRACT

This dissertation analyzes the linguistic behaviour of specific individuals in relation to different phonological phenomena, evaluating the organization of the phonological component. It is suggested that individual variability, attested in speech production, is part of linguistic representation. The following multi-representational theories are considered: Usage-based Phonology (BYBEE, 2001), Exemplar Model (PIERREHUMBERT, 2001b, 2003a) and Probabilistic linguistics (BOD; HAY; JANNEDY, 2003).

In this research, the leader concept of linguistic change as proposed by Labov (2001, v. 2) is considered. The methodological design includes two study types: a cross-sectional study of 12 individuals (6 leaders and 6 non-leaders) and a longitudinal study of 2 individuals that were selected in the cross-sectional sample. The longitudinal study is composed of 6 data collection sessions of each participant. The findings demonstrate the relevant inter and intra-individual variation and present aspects that are consistent with theoretical models adopted in this work, such as the following: lexical dynamicity and plasticity; patterns of lexical diffusion of phonological change; probabilistic behaviour of language; individual lexicon and grammar in use.

## CAPÍTULO 1

### INTRODUÇÃO

Este estudo analisa a organização do componente fonológico segundo variações no comportamento do indivíduo<sup>1</sup>, inter e intra-individuais, em relação a diferentes fenômenos fonológicos.

Variações interindividuais são aquelas observadas entre indivíduos pertencentes a um grupo relativamente homogêneo<sup>2</sup>. Um exemplo de variação interindividual seria, para um fenômeno como o cancelamento de “r” final em nominais, a observação de diferenças no comportamento lingüístico de dois indivíduos de um mesmo grupo. Assim, tomando-se como exemplo a palavra *amor*, enquanto um indivíduo cancelaria o “r” final de tal palavra, produzindo a forma reduzida *amô*, o outro não o cancelaria, de modo a produzir tipicamente a forma plena *amor*.

Já as variações intra-individuais ocorrem em um único indivíduo. No caso do cancelamento de “r” final em nominais citado, poderia haver variação intra-individual com esse mesmo falante que tipicamente não cancela esse som. Desse modo, tal indivíduo apresenta alternância entre o cancelamento e o não-cancelamento de “r” final em nominais: por exemplo, ora ele diria *amor*, ora ele diria *amô*, o que configura uma variação intra-individual.

---

<sup>1</sup> Na presente tese, como avaliamos a produção da fala e não a percepção, utilizamos o termo *indivíduo* como sinônimo de *falante*.

<sup>2</sup> Neste estudo, consideramos como indivíduos pertencentes a um grupo relativamente homogêneo aqueles que têm mesmo sexo, faixa etária, classe social, escolaridade, origem geográfica – categorias sociais propostas por Labov (1966) – e até mesma rede social – categoria social proposta por Milroy (1987).

Sugere-se que a variabilidade atestada na produção de fala dos indivíduos faz parte do conhecimento lingüístico deles (JOHNSON; MULLENIX, 1997). Com base em tal consideração, neste trabalho, adotamos teorias multirrepresentacionais, quais sejam: Fonologia de Uso (BYBEE, 2001), Teoria de Exemplares (PIERREHUMBERT, 2001b, 2003a) e Lingüística Probabilística (BOD; HAY; JANNEDY, 2003). Tais modelos contemplam a possibilidade de se considerar o indivíduo como unidade de análise.

Tradicionalmente, o comportamento do indivíduo não é considerado em trabalhos que tratam de mudança sonora. Hazen (2002, p. 501, 502), que aborda a família como unidade de análise em fenômenos de mudança, afirma<sup>3</sup>:

For some researches, the language variation patterns of individuals are idiolects<sup>4</sup>. But according to traditional variationist analysis, specifically Weinreich et al. (1968) and Labov (1989b), idiolects are not a theoretical reality. [...] In this view, the language variation grammar exists at the level of the speech community but not at the level of the individual. These assumptions prohibit a nested view of language variation whereby dialects are collections of similar idiolects, and languages are collections of similar dialects.

Pagotto (2004, p. 92), discutindo aspectos da Sociolingüística, afirma: “O sujeito falante laboviano é completamente assujeitado ao sistema heterogêneo da comunidade, que é, ao mesmo tempo, a fonte de tal sistema e o *locus* onde reside; é na intersubjetividade que o sistema se constrói e se manifesta (Cf. LABOV, 1972, p. 248-249)”. Essa citação de Pagotto nos mostra que o comportamento do indivíduo não é contemplado pela Sociolingüística. Talvez poderíamos dizer, que nesta, o indivíduo não é autônomo em

---

<sup>3</sup> As citações em língua inglesa presentes no decorrer deste texto encontram-se traduzidas no APÊNDICE A.

<sup>4</sup> Bloch (1948) introduziu na Lingüística o conceito de idioleto que seria, para ele, a variedade lingüística utilizada por um indivíduo concreto.

relação ao seu grupo social, definido em termos de fatores, como origem geográfica, sexo, idade, escolaridade e classe social. Milroy (1987, p. 132, 133) também aponta:

Labov (1966) notes considerable variation *within* social groups defined on the basis of age, class or sex, but argues convincingly that studies of the ‘idiolect’ are unlikely to reveal sociolinguistic structure as clearly as studies of the speech of whole social groups.[...] We should note here that Labov is not suggesting that individuals reveal only unstructured variations; he is showing specifically that the idiolect-based approach characteristic of earlier scholars (for example, Bloch, 1948; Hockett, 1958) is inadequate.

Nessa perspectiva tradicional da mudança sonora, a idéia de uma gramática única para a língua – a Gramática Universal, também conhecida como GU (Cf. CHOMSKY, 1997) – impede que se postule o comportamento do indivíduo como unidade de análise. Também uma outra noção, a qual é decorrente da GU, impossibilita a análise do comportamento do indivíduo: a idéia de uma representação lingüística única na gramática. Modelos multirrepresentacionais, que prevêm variação lingüística na representação abstrata, permitem a investigação das variações inter e intra-individuais. Em tais modelos, inclusive, é importante considerar o comportamento do indivíduo, porque ele contribui para a emergência e a dinamicidade dos sistemas lingüísticos. Assim, compartilhamos com a visão de Reinecke (2006, p. 68), a qual afirma:

A língua [...] é um sistema complexo dentro do qual existem formas e organizações dessas formas que estão em perpétuo desenvolvimento ou mudança. A mudança é o resultado de uma competição de forças adversárias. Essas forças são, por um lado, processos que favorecem a estabilidade do sistema, através, por exemplo, da convencionalização de formas lingüísticas; por outro lado, são processos de inovação lingüística. A noção fundamental para captar o desenvolvimento, a modificação, a inovação e renovação dentro do sistema é a noção de emergência gramatical. A localização da gramática e, portanto, da emergência, está no indivíduo e o processo pelo qual emergências nas gramáticas podem estender-se e reproduzir-se através do sistema, é a interação humana.

A respeito da incorporação da variabilidade no nível abstrato das representações mentais, Oliveira, M. (2006) afirma: “[...] a variação lingüística pode e deve ser prevista em termos abstratos, e [...] sua implementação, na fala, é sensível ao par [indivíduo – léxico]”. Com essa afirmação, Oliveira, M. aponta para o fato de que o léxico varia no nível individual.

Apesar de existirem visões contrárias à perspectiva de que o comportamento do indivíduo é uma unidade de análise plausível, pesquisas atuais atestam variações inter e intra-individuais em fenômenos de variação e mudança sonora (Cf. HUBACK, 2003; OLIVEIRA, M., 1992; OLIVEIRA GUIMARÃES, 2004; PAIVA; DUARTE, 2003; SANGSTER, 2002). Marshall (2004, p. 26) assegura que: “The practice of grouping informants to show patterns of linguistic variation is of course a valide procedure; however, much systematic individual variation is left unaccounted for by this method”. Também em relação ao comportamento do indivíduo, Milroy (1987, p. 133) afirma:

Nevertheless, as long as we do not expect to find in the language of the individual a perfect replica in miniature of his dialect grammar, there is no reason why a single speaker’s output should be viewed as unstructured and unworthy of study. In fact, Bickerton, Le Page and Gumperz [Bickerton (1975), Le Page (1968), Gumperz (1976b)] all take the individual as the basis of their descriptions of variation, principally because of the difficulty of defining corporate group boundaries. Le Page additionally is anxious to emphasize the individual’s freedom of choice to use his language variably as a means of identifying with different groups at different times.

Como se pode notar pela citação de Milroy, Le Page (1968) chega a atribuir uma razão para o comportamento variável do indivíduo. Para Le Page, o indivíduo varia sua linguagem segundo a identificação que ele tem com diferentes grupos em momentos diversos.

Docherty *et al.* (1997, p. 276) também apontam para a importância de se estudar o comportamento do indivíduo: “[...] patterns of variation (both within and between speakers) emerge as quite systematic, but are more complex and of greater magnitude than has been suggested in the phonological literature”. Como Docherty *et al.* (1997) sugerem, há variação sistemática “within and between speakers”. Estes dois tipos de variação são contemplados nesta tese: variações intra- e interindividuais.

Por todo o exposto acima, percebe-se a relevância desta pesquisa que contempla a variabilidade no comportamento do indivíduo, questão aventada nos diversos trabalhos atuais citados. O caráter das representações lingüísticas, mais especificamente fonológicas, também tem destaque em nosso estudo, uma vez que estas são avaliadas pelas teorias aqui adotadas e visto que, ao analisarmos a variabilidade individual, estaremos investigando a natureza de tais representações.

Para estudar variações inter e intra-individuais, investigamos três fenômenos fonológicos que envolvem casos de variação sonora no português brasileiro, relacionados a cancelamento consonantal<sup>5</sup>. Esses fenômenos são:

---

<sup>5</sup> Adotamos o termo “cancelamento” para denominar os três fenômenos em análise. Contudo, deve-se ressaltar que reconhecemos que tais fenômenos são casos de lenição. A lenição ou enfraquecimento consonantal ocorre quando uma consoante é produzida com um grau relativamente fraco de esforço muscular e força respiratória. De acordo com Harris (1994, p. 120), a lenição é um processo em que, tipicamente, um segmento passa por uma série de estágios, sendo o último deles o seu cancelamento. Para uma maior discussão a respeito de lenição, consultar Fontes Martins (2001, p. 34-37).

**QUADRO 1: Os 3 fenômenos em estudo**

<b>Fenômenos<sup>6</sup></b>	<b>Exemplos</b>	<b>Corpora</b>
<b>Fenômeno 1 (r #):</b> Cancelamento de “r” final em nominais	calor → calô	<b>Corpus 1:</b> Huback (2003)
<b>Fenômeno 2 (v-v):</b> Cancelamento das líquidas /l/ e /r/ intervocálicas	óculos, espírito → ócus, espíto	<b>Corpus 2:</b> Fontes Martins (2001)
<b>Fenômeno 3 (cc):</b> Cancelamento das líquidas /l/ e /r/ em encontro consonantal tautosilábico	refrigerante → refígerante exemplo → exempo	<b>Corpus 3:</b> Cristófaro Silva (2000a, b, 2003a)

Todos os três fenômenos caracterizados no QUADRO 1 são provenientes de corpora que se baseiam no dialeto de Belo Horizonte. Ainda, eles foram estudados à luz da teoria da Difusão Lexical (CHEN; WANG, 1975), segundo a qual o léxico tem um papel fundamental nos fenômenos de mudança sonora. Neste trabalho, também consideramos o modelo da Difusão Lexical, já que as teorias multirrepresentacionais aqui adotadas compartilham da noção de importância do léxico.

Tendo em vista os três fenômenos em análise, esta pesquisa tem como objetivo geral avaliar a variabilidade individual, para uma maior compreensão da natureza das representações lingüísticas – com foco nas representações fonológicas – organizadas probabilisticamente. Como objetivos específicos, esta tese busca:

1. caracterizar o comportamento do indivíduo quanto a diferentes fenômenos fonológicos do português brasileiro, mais especificamente, quanto aos três fenômenos descritos no QUADRO 1. Sangster (2002), analisando sete

<sup>6</sup> No decorrer deste trabalho, os Fenômenos 1 (r #), 2 (v-v) e 3 (cc) serão mencionados por diversas vezes. Em razão disso, a fim de facilitar a leitura, como um recurso mnemônico, adotamos os códigos (r #), (v-v) e (cc) para cada um desses fenômenos respectivamente.

- fenômenos fonológicos do inglês, encontrou variações inter e intra-individuais relevantes;
2. avaliar a variabilidade individual com relação a itens específicos, ou seja, quanto ao fator palavra, enfatizado pela Difusão Lexical e pelas teorias multirrepresentacionais em análise;
  3. caracterizar o comportamento individual com relação a diferentes estilos. Também no trabalho de Sangster (2002), o estilo foi um importante fator para a ocorrência de variações inter e intra-individuais;
  4. avaliar a proposta de Labov (2001, v. 2) quanto à questão da liderança em casos de mudança sonora. Este autor sugere que há indivíduos que são avançados (os líderes) em relação às mudanças sonoras, por apresentarem índices mais altos de realização da variável inovadora. Em contrapartida, haveria indivíduos que são conservadores (não-líderes), por apresentarem índices mais baixos dessa realização;
  5. analisar, com relação a fenômenos de mudança sonora, o comportamento de indivíduos adultos em estudo longitudinal (PAIVA; DUARTE, 2003). Em Lingüística, esse tipo de estudo é amplamente realizado em trabalhos de aquisição da linguagem com crianças (BONILHA, 2004; MIRANDA, 2007; OLIVEIRA GUIMARÃES, em andamento); contudo, não o é em trabalhos como o atual que analisa fenômenos de variação sonora em indivíduos adultos.

Esta tese é composta de mais outros seis capítulos, além deste primeiro capítulo de introdução, o que pode ser conferido a seguir:

- no segundo capítulo, é realizada uma revisão da literatura relacionada ao tema deste estudo. Nesta ocasião, serão discutidos especialmente os textos de:

Oliveira Guimarães (2004), Huback (2003), Labov (2001, v. 2), Milroy (1987), Oliveira, M. (1992), Paiva e Duarte (2003), Sangster (2002), Shen (1990);

- no terceiro capítulo, será feita uma apresentação e discussão da base teórica que norteia esta pesquisa;
- no quarto capítulo, é apresentada a metodologia empregada neste estudo. São apontados e discutidos nesta ordem: o tipo de método adotado, a amostra trabalhada, o material utilizado (de todo o aparato e estímulos) e os procedimentos de coleta e análise dos dados dos estudos transversal e longitudinal realizados;
- no quinto capítulo, é realizada a análise dos dados e discussão dos resultados do estudo transversal, por meio de um tratamento estatístico conferido aos dados;
- no sexto capítulo, além de ser apresentada uma análise de caráter longitudinal, é feita a análise dos dados e discussão dos resultados do estudo longitudinal aqui realizado, também por meio de um tratamento estatístico dos dados;
- no sétimo capítulo, são apresentadas as considerações finais deste estudo, com uma retomada dos resultados obtidos e sugestões de pesquisas futuras.

## CAPÍTULO 2

### O COMPORTAMENTO DO INDIVÍDUO

#### 2.1 Introdução

Esta tese trata do comportamento lingüístico de indivíduos específicos no exame de três fenômenos fonológicos. Sugere-se que a variabilidade atestada na produção é parte do conhecimento lingüístico dos falantes (JOHNSON; MULLENIX, 1997). Neste capítulo, trataremos da literatura relativa a esse tema. Assim, primeiramente, na seção 2.2, abordamos os estudos de Oliveira, M. (1992), Huback (2003) e Oliveira Guimarães (2004), tendo em vista que esses autores sugerem a análise do comportamento do indivíduo em relação a fenômenos de mudança e variação sonora. Na seção 2.3, discutimos o trabalho de Paiva e Duarte (2003), as quais reconhecem irregularidades no comportamento do indivíduo e apontam para a difusão gradual da mudança sonora na comunidade. Ainda em 2.3, abordamos Shen (1990) que trata desta última questão. Na seção 2.4, tratamos de três abordagens possíveis<sup>7</sup> na análise de fenômenos de mudança sonora:

- 1) a abordagem da comunidade lingüística, proposta na Sociolingüística por Labov (1966), em relação a fatores, como: classe social, escolaridade, sexo, faixa etária e origem geográfica;
- 2) a abordagem das redes sociais, proposta por Milroy (1987);

---

<sup>7</sup> Deve-se destacar que há outras propostas de abordagem para a mudança sonora na literatura, como o citado Hazen (2002), por exemplo, que aponta a *família* como unidade de análise.

- 3) a abordagem do comportamento do indivíduo, realizada por Sangster (2002) e proposta neste estudo.

Vale ressaltar que, após a discussão das duas primeiras abordagens (a da comunidade e a das redes sociais), já entrando na terceira abordagem (a do comportamento do indivíduo), trataremos da questão da liderança (LABOV, 2001, v. 2). Esta é muito importante à discussão do presente trabalho, tendo servido de base, inclusive, para a seleção dos nossos participantes (Cf. “Metodologia”, capítulo 4). Ainda, nessa terceira abordagem, tratamos especialmente de Sangster (2002), cujo estudo nos motivou a investigar o comportamento do indivíduo quanto a diferentes estilos e fenômenos fonológicos (Cf. “Metodologia”). Por fim, na seção 2.5, apresentamos um sumário deste capítulo.

## **2.2 O comportamento do indivíduo como unidade de análise**

Propomos, neste estudo, que se considere o comportamento do indivíduo como unidade de análise. Alguns trabalhos que compartilham dessa proposta e que são abordados nesta seção são: Oliveira, M. (1992), Huback (2003) e Oliveira Guimarães (2004).

Começando pelo artigo de Oliveira, M. (1992), este autor adota a Difusão Lexical como teoria. Ele tem como foco principal discutir o papel do contexto fonético nas mudanças sonoras, defendendo a idéia de que tal contexto atua *a posteriori*, como um possível estabilizador das inovações lingüísticas. Segundo Oliveira, M., o argumento de que o

contexto fonético atua *a posteriori* é compatível com a abordagem difusionista da mudança sonora.

Oliveira, M. (1992) argumenta ainda que é preciso investigar o comportamento do indivíduo em estudos sobre mudança lingüística, sugerindo que, muitas vezes, indivíduos de um único grupo apresentam comportamentos lingüísticos diferentes para um mesmo fenômeno. Oliveira, M. (1992, p. 39) assevera ainda que o indivíduo é mais homogêneo do que o grupo.

[...] não há variação a nível individual [nos dados analisados por Oliveira, M. (1992)]: o informante 11 tem as formas do verbo **jogar** na lista do [u]-categórico, enquanto que os informantes 1 e 6 as têm na lista do [o]-categórico. Numa análise variacionista clássica, tendo como perspectiva o grupo, teríamos um percentual de 25% de alçamento. Contudo, na perspectiva do indivíduo o percentual de variação é 0%! Sugiro, então, mesmo correndo o risco de heresia, que o comportamento do indivíduo é mais homogêneo do que o comportamento do grupo. (OLIVEIRA, M., 1992, p. 39).

A afirmação de Oliveira, M. demonstra que é relevante o estudo do comportamento do indivíduo que fazemos nesta tese. Porém, especialmente o trecho, “Contudo, na perspectiva do indivíduo o percentual de variação é 0%! Sugiro, então, mesmo correndo o risco de heresia, que o comportamento do indivíduo é mais homogêneo do que o comportamento do grupo”, pode ser confrontado com a idéia defendida nesta tese de que existem variações, não somente inter, mas também intra-individuais.

Enquanto Oliveira, M. (1992) contempla apenas as variações interindividuais, a presente tese considera que existem também variações intra-individuais, partindo da premissa de que um único indivíduo pode ter um comportamento lingüístico heterogêneo. Vale acrescentar que propomos que o comportamento heterogêneo do indivíduo é regulado por parâmetros probabilísticos, gerenciados pelo uso lingüístico e

refletindo a linguagem como um fenômeno social. Essa proposta será desenvolvida na apresentação e discussão dos resultados, nos capítulos 5 e 6. Concluindo o trabalho, Oliveira, M. (1992, p. 40) afirma ainda:

A menos que haja alguma razão séria em contrário, sugiro que o comportamento do indivíduo seja checado para todos os itens lexicais. Somente depois disto é que os indivíduos podem ser agrupados, se isto for possível.

Essa afirmação de Oliveira, M. aponta para o fato de que o comportamento do indivíduo deve ser considerado como unidade de análise em trabalhos sobre mudança e variação sonora. Contudo, como se mencionou acima, Oliveira, M. considera apenas a variação interindividual, ao passo que esta tese, além de considerar tal tipo de variação, considera ainda a intra-individual.

Huback (2003) também aponta o fator indivíduo como relevante em seu trabalho. A autora analisa, com base na teoria da Difusão Lexical, o fenômeno do cancelamento de “r” final em nominais, o qual é um dos três fenômenos analisados neste estudo (Cf. QUADRO 1, capítulo 1). Huback testou, por meio do programa de análise estatística Goldvarb, uma série de fatores, dentre os quais o programa selecionou, como relevante, o fator indivíduo (Cf. HUBACK, 2003, p. 105). A autora reconhece que pode haver variação no nível do comportamento do indivíduo, quando afirma que: “Não podemos esquecer, contudo, que, embora pertença a um grupo específico, cada falante é um indivíduo e, como tal, pode apresentar um comportamento divergente do grupo a que pertence” (HUBACK, 2003, p. 114). Essa afirmação de Huback aponta mais especificamente para a variabilidade interindividual.

Duas afirmações dessa autora mostram exemplos de variação no comportamento do indivíduo (variação interindividual), ocorridos em sua pesquisa. Esta primeira afirmação diz respeito aos informantes favorecedores do cancelamento do “r” final:

Observe-se, por exemplo, o comportamento das informantes E e Z. Ambas pertencem à classe média, à primeira faixa etária, ao mesmo sexo e ao nível superior de escolaridade. No entanto, o cancelamento do (r) nas duas informantes é completamente diferente. A informante E apresenta probabilidade de cancelamento de 0,759, enquanto a informante Z, de 0,182. A probabilidade da informante Z é tão baixa quanto a da informante N, que é da terceira faixa etária, também mulher, de classe média e com nível superior de escolaridade. (HUBACK, 2003, p. 114).

Já esta segunda afirmação se refere aos indivíduos que mais preservam o “r” final:

A análise do perfil dos informantes que mais preservam o (r) corrobora aquilo que foi dito anteriormente sobre o falante como um indivíduo com comportamento lingüístico heterogêneo dentro de seu grupo. [...] o que se percebe, na verdade, é que há falantes que não se conformam ao padrão do grupo a que pertencem, apresentando um comportamento atípico, o que demonstra que essa padronização não pode ser categoricamente afirmada. (HUBACK, 2003, p. 115).

Huback (2003, p. 116) continua:

O que estou pretendendo concluir [...] é que, entre indivíduos com o mesmo perfil, não há um comportamento único no que concerne ao cancelamento do (r) final em nominais. Não são, portanto, dados relativos a idade, sexo, escolaridade ou classe social que, isoladamente, determinarão se o indivíduo será mais ou menos propenso ao cancelamento do (r). Existem, possivelmente, outros fatores que interferem nessa questão.

Como se pode perceber por meio desta última citação, Huback (2003) realmente atesta variabilidade no nível do comportamento do indivíduo. Contudo, a autora não aprofunda nessa questão, apenas aponta que “existem, possivelmente, outros fatores que interferem” no comportamento do indivíduo, sem se deter neles. Esta tese pretende avançar na formulação de Huback (2003), apontando variáveis que possam interferir no comportamento lingüístico individual dos falantes. Argumentamos que fatores como

*fenômeno, estilo e palavra* podem ser importantes para a compreensão da variabilidade individual, por isso tratamos tais fatores como variáveis independentes nesta pesquisa (Cf. capítulo 4).

Outro trabalho que vale ser destacado neste estudo é Oliveira Guimarães (2004), a qual analisa a variação sonora nas seqüências de (sibilante + africada alveopalatal), no português atual falado em Belo Horizonte, com base nas teorias da Difusão Lexical, da Fonologia de Uso e do Modelo de Exemplares. Esta autora, como Huback (2003), submeteu os seus dados ao programa de análise estatística Varbrul que selecionou o fator não-estrutural indivíduo no caso do fenômeno de cancelamento da africada. Em sua análise, Oliveira Guimarães (2004, p. 73) aponta:

A análise da variável indivíduo mostra que os informantes que mais cancelam são os informantes n e c. O informante f é o que menos cancela, seguido pelos informantes d, h, i e p. É importante lembrar que este corpus possui um grupo de informantes bastante homogêneo. Além de serem naturais de Belo Horizonte, terem o mesmo grau de escolaridade e estarem dentro da mesma faixa etária, todos os informantes estudam na mesma faculdade. [...] Assim, mesmo em um grupo homogêneo como esse, identificam-se indivíduos que destoam do resultado geral apresentado.

Como se pode notar por essa citação de Oliveira Guimarães, em sua pesquisa, foi atestada a relevância do fator comportamento do indivíduo. Essa autora, como Oliveira, M. (1992) e Huback (2003), dá destaque à variação interindividual, ao defender que em um grupo homogêneo de sua análise, há indivíduos que “destoam” deste. Como frisamos, esta tese analisa não somente as variações interindividuais, mas as intra-individuais também.

Oliveira Guimarães sugere um aprofundamento no estudo do comportamento do indivíduo, quando demonstra os resultados de sua pesquisa: “Questões relacionadas aos indivíduos poderiam oferecer pistas para que se possa compreender melhor esse

resultado” (OLIVEIRA GUIMARÃES, 2004, p. 67, 68). A autora argumenta que “[...] no estudo da variação e mudança sonora, é interessante que se analise o falante individualmente. A lista de itens lexicais que sofrem a mudança em determinado tempo pode não ser a mesma para cada indivíduo” (p. 32). Com esta afirmação, a autora, além de sugerir que se considere o fator comportamento do indivíduo, oferece indícios para uma importante idéia defendida nesta tese: a de que indivíduos diferentes, no uso, podem ter léxicos e, conseqüentemente, gramáticas e representações lingüísticas diferentes (BYBEE, 2001). Vale ressaltar que há comunicabilidade entre os indivíduos, a qual é decorrente do uso em comunidade (REINECKE, 2006). Oliveira Guimarães afirma ainda:

Infelizmente, não se dispõe de mais informações sobre esses indivíduos, como: bairro onde residem, escolaridade e profissão dos pais, etc. Tais informações poderiam contribuir para traçar o perfil dos indivíduos inovadores e dos indivíduos mais resistentes ao cancelamento da africada. (OLIVEIRA GUIMARÃES, 2004, p. 73).

Por essa última citação, percebe-se que Oliveira Guimarães não teve a possibilidade de investigar, mais profundamente, o comportamento do indivíduo, sendo que ela cita motivações sociais para a variabilidade individual. Neste estudo, como dissemos, destacamos fatores, como *fenômeno*, *palavra* e *estilo*, como possíveis variáveis dessa variabilidade. É interessante notar ainda que, nessa citação, a autora aponta para a questão da liderança (LABOV, 2001, v. 2), a ser discutida na próxima seção, ao fazer menção a indivíduos inovadores (líderes) e mais resistentes (não-líderes) em relação ao fenômeno sonoro por ela analisado.

Os textos discutidos ao longo desta seção demonstram que a variabilidade individual, com destaque para a variação interindividual, merece ser investigada. Contudo, tais textos sugerem, mas não chegam a adotar o indivíduo como unidade de análise. Este

trabalho se propõe a fazer isto, abordando não somente as variações interindividuais, mas também as intra-individuais.

Na próxima seção, tratamos dos textos de Paiva e Duarte (2003) e Shen (1990), os quais apontam para a difusão gradual da mudança sonora na comunidade, ou seja, para o fato de que a mudança sonora atinge falante a falante, sendo que, com Paiva e Duarte (2003), abordamos também “irregularidades no comportamento do indivíduo”.

### **2.3 Irregularidades no comportamento do indivíduo e difusão gradual da mudança sonora na comunidade**

Paiva e Duarte (2003) é um trabalho que merece ser destacado neste estudo, porque sinaliza a importância de se considerar o comportamento do indivíduo, quando aborda irregularidades nesse comportamento e, também, quando aborda a idéia da difusão gradual da mudança sonora na comunidade. Também Shen (1990) merece ser aqui citado, por reconhecer essa difusão gradual da mudança na comunidade.

O trabalho de Paiva e Duarte (2003) reuniu artigos de pesquisadores do Programa de Estudos sobre o Uso da Língua (PEUL), que realizaram estudos na cidade do Rio de Janeiro. Os pesquisadores do PEUL investigaram diversos fenômenos de variação fonético-fonológica (monotongação de ditongos, palatalização e outros) e variação morfossintática (concordância nominal e verbal, uso do artigo, etc.). Os seus dados provêm de duas coletas em dois tempos diferentes. Na primeira coleta, havia 64 informantes. Na segunda, com um intervalo de tempo em relação à primeira entrevista

variando entre 16 e 19 anos, havia 16 informantes. Como um resultado geral encontrado em quase todos os artigos, Paiva e Duarte (2003, p. 28) apontam:

Em vários dos trabalhos [apresentados no livro] destaca-se, no entanto, a irregularidade no comportamento dos indivíduos, quando comparados entre si: alguns indivíduos apresentam nítida estabilização de comportamento lingüístico no período de tempo considerado, reforçando, assim, a hipótese clássica acerca da fixação do sistema; outros apresentam flutuações quantitativas ou qualitativas ao longo dos anos que separam as duas amostras de fala, demonstrando, assim, algumas descontinuidades de comportamento em função da mudança de faixa etária, do nível de escolarização ou da imposição de exigências do mercado lingüístico, embora nem sempre seja fácil isolar a variável responsável por essas alterações.

Como se nota por essa afirmação, as autoras apontam que indivíduos podem alterar ou não o seu comportamento lingüístico em função do tempo. Em nossa pesquisa, avaliaremos essa afirmação de Paiva e Duarte (2003, p. 28) por meio dos estudos de caráter longitudinal que realizamos (Cf. capítulo 6). Ainda, podemos perceber por essa citação que, quanto aos indivíduos que alteraram seu comportamento, as autoras indicam alguns fatores que potencialmente teriam causado “irregularidade no comportamento” deles, como “mudança de faixa etária, do nível de escolarização”. Vale destacar que as “irregularidades” de que tratam Paiva e Duarte (2003, p. 28) se referem somente à variação intra-individual. Esta, por sua vez, é resultante do fator tempo apenas, já que a variabilidade encontrada pelas autoras no nível individual se relaciona com os dois momentos distintos de coleta de dados. A presente pesquisa deve avaliar também outros fatores de variação tanto inter quanto intra-individual, como o estilo de fala, por exemplo, sugerido por Sangster (2002).

O artigo de Paiva e Duarte (2003) é interessante ainda a este trabalho, porque além de reconhecer a variabilidade individual (em especial a intra-individual até este momento), aponta para o fato de que a mudança difunde-se gradualmente na comunidade. Assim,

Paiva e Duarte (2003, p. 13) comentam: “A atividade humana da linguagem caracteriza-se por um conflito entre duas faces aparentemente contraditórias: de um lado, uma aparência de estabilidade e, de outro, a constante variação e mudança tanto no indivíduo como na comunidade”. Como se vê, essas autoras reconhecem a variação e a mudança no nível do indivíduo. Ao dizerem que existe na língua uma “constante variação e mudança tanto no indivíduo como na comunidade”, Paiva e Duarte apontam para o fato de haver variação ou representações lingüísticas múltiplas no nível individual.

Essas autoras consideram ainda que a forma como a mudança atinge a comunidade é gradual: “Em situação de mudança lingüística, instaura-se uma luta evolutiva [...] entre as novas formas que passaram a concorrer com as formas antigas, com as primeiras se espalhando de um falante para outro” (PAIVA; DUARTE, 2003, p. 15). As autoras mostram, desse modo, que a forma inovadora não atinge todos os falantes de uma só vez: ela se espalha de falante para falante, o que gera variações inter e intra-individuais. Esta última afirmação de Paiva e Duarte reforça, assim, a idéia de que a mudança sonora difunde-se gradualmente na comunidade.

Também Shen (1990, *apud* OLIVEIRA GUIMARÃES, 2004, p. 32) aponta para a difusão gradual da mudança sonora na comunidade quando afirma que a mudança pode ocorrer de modo distinto de um falante para outro:

[...] it is very likely that a different word will change its phonological category from A to B at a quite different time by its learners. What is observed is that lexical irregularity appears. Along the social dimension, because of various language environments, while one individual has changed the phonological identity of a word, at the same time many others may still keep the original one. (SHEN, 1990, p. 166).

A fim de aprofundarmos a discussão sobre o comportamento do indivíduo e tendo em vista que trabalhamos com fenômenos que são casos de variação sonora (Cf. QUADRO 1), na próxima seção, tratamos de três abordagens possíveis para fenômenos de mudança e variação sonora citados na introdução deste capítulo: 1) a abordagem da comunidade de fala (LABOV, 1966); 2) a abordagem das redes sociais (MILROY, 1987); e 3) a abordagem do comportamento do indivíduo (SANGSTER, 2002). Como mencionado (Cf. seção 2.1), na análise das abordagens (1) e (2), trataremos da questão da liderança (LABOV, 2001, v. 2).

#### **2.4 Abordagens para fenômenos de mudança e variação sonora: comunidade lingüística, redes sociais e comportamento do indivíduo**

Neste estudo, consideramos três abordagens possíveis para fenômenos de mudança e variação sonora. Assim, nesta seção, tratamos primeiramente de duas abordagens *supra-individuais* (PAGOTTO, 2004, p. 75): a da comunidade lingüística (LABOV, 1966) e a das redes sociais (MILROY, 1987). Nesse ponto, já passando para a terceira abordagem, a do comportamento do indivíduo (SANGSTER, 2002), discutimos a questão da liderança (LABOV, 2001, v. 2), considerada no presente trabalho. Note-se que, na discussão dessas três abordagens, há um percurso que parte de uma unidade maior (comunidade lingüística), passando uma unidade intermediária (redes sociais), até chegar a uma unidade menor (indivíduo).

Começamos a tratar da primeira abordagem, a da comunidade lingüística. Labov (1966) propõe a Sociolingüística, segundo a qual, na análise da mudança ou variação

lingüística, fatores lingüísticos devem ser correlacionados a fatores sociais da comunidade lingüística, tais como classe social, escolaridade, sexo, faixa etária e origem geográfica. Vale destacar que, dentre tais fatores, o autor enfatiza a estratificação da sociedade em classes. De acordo com Ash (2002, p. 402), “Social class is a central concept in sociolinguistic research [...]”. De fato, o que vemos é que, na Sociolingüística, a comunidade de fala se organiza, em primazia, por meio de classes sociais. Contudo, vale destacar que o conceito de classe social é questionável, uma vez que não é universal. Muitas sociedades, como as tribais de forma geral, não se estruturam em termos de classes sociais.

A respeito da estruturação da comunidade lingüística em classes sociais, Oliveira, M. (1992, p. 40) pondera: “A divisão da comunidade de fala em grupos (ou classes, ou estratos) não é, necessariamente, o último estágio da divisão. Na verdade, ninguém demonstrou ainda que esta fosse a divisão correta.” Por essa afirmação de Oliveira, M., percebemos que o autor questiona a divisão clássica que a Sociolingüística faz da comunidade de fala. Também Pagotto (2004, p. 71), tratando a respeito desta, afirma: “Um dos conceitos mais problemáticos da Sociolingüística é o de *comunidade lingüística*” e questiona (Cf. PAGOTTO, 2004, p. 49):

Qual é o papel da dimensão subjetiva no funcionamento da variação, se noções como *idioleto* e a fala do indivíduo não são, como aponta Labov (1972a) a fonte da variação, mas a comunidade lingüística? Ao mesmo tempo, como lidar com fenômenos de variação em que a dimensão subjetiva parece desempenhar um papel crucial [...]?

Nesse questionamento de Pagotto a respeito da comunidade lingüística de Labov<sup>8</sup>, vemos que o autor chega a apontar a dimensão subjetiva, o indivíduo, como relevante na análise de fenômenos de variação.

Passemos à segunda abordagem, a das redes sociais, que é proposta por Milroy (1987) em contrapartida a Labov (1966). A autora considerou insuficiente a análise sociolingüística de estratificação em classes da comunidade lingüística, proposta por este autor.

Em seu estudo, Milroy observou que, por exemplo, dois indivíduos de mesmo sexo, faixa etária, escolaridade, origem geográfica e classe social (fatores sociais condicionadores da mudança e variação lingüística, segundo a Sociolingüística) apresentavam comportamentos diferentes em relação a um único fenômeno lingüístico<sup>9</sup>. Isso é apontado em: “Yet, despite these linguistic differences which can be linked firmly to the variables of sex, age and area, there is a large residue of systematic variation between individuals which cannot be characterized in any clear way by dividing speakers into further subgroups” (MILROY, 1987, p. 131). É a partir dessa idéia da variação sistemática entre indivíduos que Milroy propõe um outro tipo de agrupamento social dos indivíduos: as redes sociais. Na proposta da autora, estas se caracterizariam por representarem as relações sociais imediatas de um indivíduo, por exemplo, no seu trabalho, na sua família, na sua vizinhança, na sua escola, etc. Assim, as redes ligariam indivíduos diferentes, mas com especificidades nas relações, atentando-se para o fato de que a fala do indivíduo é distinta, de acordo com o interlocutor. Por exemplo, um

---

<sup>8</sup> Uma discussão mais ampla a respeito da noção de *comunidade lingüística*, de Labov (1972) é feita por Pagotto (2004, p. 71-77).

<sup>9</sup> Vale lembrar que os trabalhos de Oliveira, M. (1992), Huback (2003) e Oliveira Guimarães (2004), discutidos na seção anterior, verificaram, assim como Milroy (1987), que há diferença no comportamento de indivíduos de um grupo relativamente homogêneo.

indivíduo, na sua relação específica de trabalho, conversa de uma forma com sua chefe, que é diferente da forma como conversa com seu irmão na relação familiar.

Milroy (1987, p. 135) afirma: “[...] the network concept was developed for the very purpose for which we require it now – to explain individual behaviour kinds which cannot be accounted for in terms of corporate group membership”. Dessa forma, a noção de redes sociais seria uma alternativa para explicar a variação lingüística, no lugar da noção de estratificação social em classes (Cf. MARSHALL, 2004, p. 23). Inclusive, Milroy aponta a dificuldade de se propor essa estratificação na sociedade, por nem sempre ser possível delimitar claramente as suas classes sociais (Cf. MARSHALL, 2004, p. 23).

É oportuno agora tratar da questão da liderança que é considerada neste estudo (Cf. capítulo 4). Labov (2001, v. 2) retoma o trabalho de Milroy (1987), a fim de discutir as redes sociais. Ele aponta que nessas redes há líderes e não-líderes da mudança lingüística, mas sua análise se concentra nos líderes. Na proposta desse autor, a classificação de um falante como líder ou não-líder se baseia nos índices estatísticos obtidos para esse falante, em relação a uma variável inovadora que reflete uma mudança lingüística. Assim, um líder apresenta maiores índices de realização de uma variável inovadora do que um falante não-líder que apresenta o inverso, ou seja, menores índices de realização dessa variável<sup>10</sup>.

Labov (2001, v. 2) afirma que, conhecendo os líderes da mudança, de certa forma, é possível conhecer as causas ou as motivações desta. Assim, o autor se detém na análise dos líderes, investigando o papel e o perfil deles. Considerando as redes sociais, Labov

---

<sup>10</sup> Deve-se destacar que, nesta tese, assumimos essas noções de líder e não-líder, de Labov, na análise dos três fenômenos aqui avaliados (Cf. “Metodologia”, capítulo 4).

concluiu a respeito dos líderes que eles geralmente têm muitos contatos sociais dentro e fora do seu grupo local, o que ele chama de “centralidade expandida” (LABOV, 2001, v. 2, p. 364). Segundo o autor, os líderes estão localizados no centro da hierarquia social e têm contato com pessoas de diferentes classes: tanto se relacionam com pessoas de classes baixas quanto de classes altas. Todas essas características dos líderes contribuiriam para o fato de eles estarem à frente no uso de uma variável inovadora. Ainda, nas mudanças em progresso, Labov observou que as mulheres geralmente são líderes. De acordo com o autor, os líderes também têm uma história de vida de não-conformidade social: “The history of our leaders of linguistic change is a history of nonconformity, and their sociolinguistic position is a display of nonconformity” (LABOV, 2001, v. 2, p. 410). Relacionado a isso, Labov (2001, v. 2, p. 410) afirma: “[...] negotiation, persuasion, and denunciation, all enlisted under a profound intolerance for cuspidity, hypocrisy, and injustice. These are the qualities that make a great leader of linguistic change”<sup>11</sup>.

Neste momento, entramos na terceira abordagem, a do comportamento do indivíduo, que pode ser articulada à questão da liderança proposta por Labov (2001, v. 2). A existência de indivíduos líderes e não-líderes é uma evidência para a variabilidade individual nos casos de mudança sonora. Além disso, é uma evidência para a difusão gradual da mudança sonora na comunidade (SHEN, 1990; PAIVA; DUARTE, 2003). Afinal, se a mudança atingisse toda a comunidade de uma só vez, não se esperaria que houvesse indivíduos líderes e não-líderes. Desse modo, a questão da liderança merece ser considerada neste estudo por iluminar aspectos relacionados à variabilidade individual.

---

<sup>11</sup> Deve-se destacar esta pesquisa pretende contribuir para o delineamento das propriedades dos indivíduos líderes e não-líderes (LABOV, 2001, v. 2). Contudo, analisar os correlatos sociais apresentados nesse parágrafo, na definição do perfil dos líderes da mudança lingüística, vai além dos propósitos desta tese. Seria interessante que pesquisas posteriores fizessem esse estudo.

Ainda com relação à terceira abordagem do comportamento do indivíduo, destacamos um fato que contraria as propostas de Labov (1966) e Milroy (1987), relativas à primeira e à segunda abordagem respectivamente: observa-se que indivíduos com mesmo sexo, faixa etária, classe social, origem geográfica e escolaridade, e até pertencentes a uma mesma rede social, apresentam comportamentos diferentes em relação a um determinado fenômeno lingüístico. Tal fato aponta para a importância de se analisarem as variações interindividuais especialmente. Vimos, na seção anterior, que autores como Oliveira, M. (1992), Huback (2003) e Oliveira Guimarães (2004) já destacam que analisar o comportamento do indivíduo em fenômenos de mudança e variação sonora é fundamental. É possível, ainda, que um único indivíduo apresente um comportamento variável em relação a um dado fenômeno, o que caracteriza a variação intra-individual que aqui também investigamos. Pagotto (2004, p. 90), discutindo a análise sociolingüística, afirma:

Muitas vezes, um sujeito *fala* com uma voz que, aos olhos do pesquisador, não é a sua: não é sua idade, não é sua faixa etária, não é seu grau de escolaridade. Essas falhas costumam ser descartadas, quando são justamente o índice mais premente do processo de identidade [...].

Considerando essa afirmação de Pagotto, vemos que as “falhas” encontradas em uma análise sociolingüística apontam para a importância de se considerar a variabilidade individual. Sangster (2002) também discute o comportamento do indivíduo, investigando variações inter e intra-individuais por meio da análise de 7 fenômenos fonológicos no inglês de Liverpool. Para tanto, a autora fez experimentos com análise acústica dos dados (experimentos sócio-fonéticos) e entrevistas, por meio das quais buscou incorporar as percepções e opiniões dos falantes de seu estudo.

É interessante mencionar que um dos experimentos de Sangster mostrou que atributos individuais dos falantes – tais como as suas redes sociais e os seus planos para o futuro – são tão importantes na variação quanto o *status* sócio-econômico deles. Como um dos resultados de sua pesquisa, Sangster observou que os falantes variaram a sua pronúncia segundo o estilo de fala. Com base nesse achado da autora, propomo-nos a avaliar o fator estilo nesta pesquisa (Cf. capítulo 4, “Metodologia”).

Sangster notou também que seus informantes não mostraram os mesmos padrões de variação para todas as sete variáveis fonéticas investigadas. Assim, um falante que se mostrou inovador em relação a um determinado fenômeno fonológico, comportou-se de forma conservadora em relação a um outro fenômeno. Esse é um outro aspecto interessante a ser avaliado pela presente pesquisa, constituindo um de nossos objetivos: caracterizar o comportamento do indivíduo quanto a diferentes fenômenos fonológicos do português brasileiro, mais especificamente, quanto aos três fenômenos em estudo (Cf. QUADRO 1, capítulo 1).

Vale destacar que esse resultado de Sangster problematiza a noção de líder proposta por Labov. Depreende-se da leitura de Labov (2001, v. 2) que um líder apresenta esse perfil com relação a todos os fenômenos fonológicos, porque o autor se detém nestes. Já o que vemos no trabalho de Sangster é o contrário: o perfil de líder pode ser dinâmico, visto que o estudo desta autora revelou que um falante que foi inovador em relação a um fenômeno fonológico foi conservador em relação a um outro fenômeno. Neste trabalho, investigamos três diferentes fenômenos fonológicos, também com a intenção de verificar se há dinamicidade ou não no perfil do líder.

Ainda sobre o trabalho de Sangster, esta autora afirma que os indivíduos variaram sua pronúncia de acordo com a audiência (público) e com o tópico. Ela diz também que falantes com maiores aspirações sociais variaram a pronúncia de certas variáveis fonéticas mais do que falantes com menos aspirações sociais e que, em seus dados, as mulheres variaram mais do que os homens.

Essas afirmações de Sangster (2002) merecem ser consideradas na presente pesquisa. Chamam à atenção as afirmações finais que a autora faz sobre os falantes de maior ambição e sobre as mulheres, que foram os informantes que mais variaram nos dados. Isso deve ser avaliado neste trabalho, em especial, com relação ao que Labov (2001, v. 2) aborda sobre os indivíduos líderes da mudança, o que foi apresentado ainda nesta seção.

Como uma reflexão geral sobre as três abordagens tratadas nesta seção, destacamos que não se pode ignorar que os conceitos de classe social (LABOV, 1966) e de redes sociais (MILROY, 1987) esclarecem muitos fatos sobre a organização da linguagem em comunidade de fala. Outros fatores da Sociolinguística, como sexo, faixa etária, escolaridade e origem geográfica, também o fazem. Contudo, como destacamos, é possível que tais conceitos não se apliquem sempre. Ainda, deve-se ressaltar que, tanto na análise sociolinguística de Labov (1966) quanto na de redes sociais de Milroy (1987), mesmo que não explicitamente, há a concepção de uma representação abstrata única. A variabilidade observada seria externa ao sistema ou conhecimento linguístico, fazendo parte do desempenho ou comportamento efetivo no uso da linguagem.

Na realidade, a questão envolvida neste caso é relativa ao caráter das representações mentais, o que foi abordado na fonologia primeiramente por Sapir (1933, In: DASCAL,

1981). Johnson e Mullenix (1997) resumem bem essa questão, ao afirmarem que há duas perspectivas principais sobre ela:

- 1) na primeira perspectiva, que é típica de teorias lingüísticas tradicionais, as representações lingüísticas são consideradas como simples e o mapeamento entre as formas abstratas e as formas atestadas nas línguas naturais é considerado como complexo. Assim, essas representações excluiriam informações redundantes – ou seja, a variação –, operando com categorias discretas;
- 2) ao contrário da primeira, nesta segunda perspectiva, que é típica de teorias multirrepresentacionais, as representações lingüísticas são consideradas como complexas e o mapeamento entre as formas abstratas e as formas atestadas nas línguas naturais é considerado como simples. Desse modo, estas representações incluiriam informações redundantes, operando com categorias gradientes.

Johnson e Mullenix (1997) se detêm na percepção e não na produção que é o foco deste estudo. Mesmo assim, o trabalho desses autores é importante aqui, tendo em vista que esta tese investigará a produção lingüística de indivíduos na segunda perspectiva citada por Johnson e Mullenix (1997). Modelos nessa segunda linha de análise, os quais adotamos, serão discutidos no próximo capítulo, quais sejam: Fonologia de Uso (BYBEE, 2001), Teoria de Exemplares (PIERREHUMBERT, 2001b)<sup>12</sup> e Lingüística Probabilística (BOD; HAY; JANNEDY, 2003).

---

<sup>12</sup> Vale ressaltar que Pierrehumbert (2001b) expande a proposta de Johnson; Mullenix (1997), tratando da produção de fala.

## 2.5 Sumário

Neste capítulo, primeiramente abordamos os textos de Oliveira, M. (1992), Huback (2003) e Oliveira Guimarães (2004), tendo em vista que eles sugerem a análise do comportamento do indivíduo em relação a fenômenos de mudança e variação sonora, apesar de não adotarem o indivíduo como unidade de análise. Como ressaltamos, tais textos demonstram a existência de variabilidade individual, com destaque para a variação interindividual.

Em seguida, discutimos os trabalhos de Paiva e Duarte (2003) e Shen (1990), os quais apontam para a difusão gradual da mudança sonora na comunidade, ou seja, para o fato de que a forma inovadora não atinge todos os falantes de uma só vez, mas falante a falante, o que gera variações inter e intra-individuais. Nesse ponto, destacamos que Paiva e Duarte (2003), ao reconhecerem irregularidades no comportamento do indivíduo, enfatizam as variações intra-individuais.

Por último, tratamos de três abordagens possíveis na análise de fenômenos de mudança sonora. Primeiramente, tratamos de duas abordagens *supra-individuais* (PAGOTTO, 2004, p. 75): a da comunidade lingüística (LABOV, 1966) e a das redes sociais (MILROY, 1987). Nesse momento, já passando para a terceira abordagem, a do comportamento do indivíduo (SANGSTER, 2002), discutimos a questão da liderança (LABOV, 2001, v. 2). Como se apontou, na discussão dessas três abordagens, há um percurso que parte de uma unidade maior (comunidade lingüística), passando uma unidade média (redes sociais), até chegar a uma unidade menor (indivíduo).

Nesse último ponto do capítulo, problematizamos o conceito de comunidade de fala (com destaque para a estratificação em classes sociais) de Labov e apresentamos o conceito de redes sociais de Milroy, que é formulado em contrapartida ao conceito de comunidade. Já na terceira abordagem, a do comportamento do indivíduo, apresentamos a proposta de Labov relativa aos líderes da mudança lingüística, proposta esta que, de certa forma, discute o citado conceito de redes sociais. Como destacamos, a noção de líder (e de não-líder) da mudança é considerada em nossa análise. Ainda nessa terceira abordagem, enfatizamos o trabalho de Sangster (2002), a qual encontra variações inter e intra-individuais, decorrentes do fator estilo, na análise de diferentes fenômenos fonológicos. Como apontamos, com base nos achados de tal autora, propomo-nos a investigar os indivíduos deste estudo quanto a diferentes estilos e fenômenos fonológicos.

O próximo capítulo apresenta e discute o suporte teórico desta tese: a Fonologia de Uso (BYBEE, 2001), a Teoria de Exemplos (PIERREHUMBERT, 2001b) e a Lingüística Probabilística (BOD; HAY; JANNEDY, 2003), que se enquadram na segunda perspectiva, multirrepresentacional, abordada por Johnson e Mullenix (1997).

## CAPÍTULO 3

### FUNDAMENTAÇÃO TEÓRICA

#### 3.1 Introdução

Este capítulo apresenta e discute as teorias que constituem a base teórica da presente pesquisa<sup>13</sup>, as quais são abordadas nesta ordem:

- 1) Fonologia de Uso (BYBEE, 2001) → seção 3.2;
- 2) Teoria de Exemplares (PIERREHUMBERT, 2001b) → seção 3.3;
- 3) Lingüística Probabilística (BOD; HAY; JANNEDY, 2003) → seção 3.4.

Por fim, a seção 3.5 apresenta um sumário do capítulo.

#### 3.2 Fonologia de Uso

Bybee (2001) propõe a Fonologia de Uso, afirmando no início de seu livro: “This book introduces into the traditional study of phonology the notion that language use plays a role in shaping the form and content of sound systems” (BYBEE, 2001, p. 1).

---

<sup>13</sup> Como se discutiu no capítulo anterior (seção 2.4), tais modelos teóricos multirrepresentacionais consideram que as representações lingüísticas são complexas (por conterem informações redundantes) e que o mapeamento entre as formas abstratas e as formas atestadas nas línguas naturais é simples (JOHNSON; MULLENIX, 1997).

Essencialmente, essa teoria postula que a experiência lingüística – semelhante a qualquer outro tipo de experiência humana – é fundamental na organização lingüística e, conseqüentemente, na organização do componente fonológico. Bybee aponta que é comum em lingüística separar fonologia de sintaxe, sincronia de diacronia e linguagem de criança da de adulto. A teoria de linguagem que ela propõe associa essas áreas, visto que considera a língua como um fenômeno complexo que só pode ser entendido, levando-se em conta a interação de todos os seus componentes com o uso. Dessa maneira, a autora afirma que sua teoria, ao contrário das teorias fonológicas anteriores, se detém não somente na estrutura da língua, mas em todas as suas interações e usos sociais (Cf. BYBEE, 2001, p. 2). Esse aspecto teórico é muito relevante para o presente estudo, tendo em vista que investigamos o comportamento do indivíduo, uma unidade de análise social e não lingüístico, de modo a considerar a língua como um fenômeno social (SANGSTER, 2002; NARO, 2003; PAIVA; DUARTE, 2003).

O modelo de Bybee postula uma confluência entre o estudo da língua e outras ciências, como a biologia e a psicologia. Para Bybee (2001, p. 17), processos cognitivos que comandam a língua não são específicos da língua, mas são os mesmos que governam outros aspectos do comportamento social e da cognição humana. Essa visão holística da autora é bem diferente da de Chomsky (1965), por exemplo, cuja proposição é analisar a língua em isolamento, como um mecanismo inato.

Dentro do modelo de Bybee, o componente fonológico é intimamente relacionado ao léxico que, por meio do significado, constitui a gramática. O léxico, por sua vez, é maximizado na medida em que apresenta alofones e outras informações específicas, como sílaba, acento, posição (JOHNSON; MULLENIX, 1997). O detalhe fonético é generalizado em representações múltiplas que se organizam probabilisticamente. Esse

ponto é muito importante a este trabalho, tendo em vista que a noção de representações múltiplas permite incorporar ao formalismo as diversas representações observadas nas variações inter e intra-individuais.

Segundo Bybee (2001), o seu modelo incorpora a redundância – já que considera que o detalhe fonético está presente no léxico mental –, a qual é compatível com o cérebro humano. A autora lembra que a redundância nas representações é essencial para a categorização (ROSCH, 1975, ROSCH *et al.*, 1976). A incorporação da redundância é possível devido à dinamicidade e à plasticidade da memória (JOHNSON, 1997). Essas considerações da Fonologia de Uso são bem diferentes das do gerativismo e de outros modelos tradicionais, os quais propõem representações simples que minimizam o léxico e operam com mapeamento complexo.

Bybee (2001, p. 6-8) apresenta os princípios básicos de um modelo fonológico de uso.

Dentre eles, temos:

- a) a experiência afeta a representação na memória → a autora afirma que palavras mais frequentes têm representação mais forte e são mais facilmente acessadas do que palavras menos frequentes. Ela tem como hipótese que o armazenamento de palavras “velhas” ajuda no conhecimento de palavras “novas”;
- b) as palavras são armazenadas como unidades inteiras no léxico → Bybee, inclusive, diz que as crianças aprendem seqüências fonológicas como partes de palavras e nunca independente delas;

- c) a categorização é baseada em similaridades → a relação morfológica emerge das similaridades semânticas e fonéticas. A categorização também organiza a armazenagem.

O que Bybee aponta em (a) é, realmente, um grande pilar da Fonologia de Uso. O fato de a experiência afetar a representação tem ligação direta com a questão da frequência, o que é discutido ainda nesta seção. Como se vê em (b), Bybee assume a palavra como unidade de armazenamento e manipulação lingüística<sup>14</sup> – e não o morfema, como propõem o estruturalismo e o gerativismo, teorias tradicionais. Para essa autora, ao entrar no léxico, a palavra não é analisada, por exemplo, em morfemas, como também não passa por regras simbólicas. Nesse caso, não existiria um componente morfológico separado e nem uma lista de morfemas, mas sim, as palavras se articulariam em redes, como será visto a seguir.

O que Bybee diz em (c) se relaciona com a sua afirmação de que a estrutura de uma palavra emerge de conexões em rede que ela mantém com outras palavras do léxico e com outros níveis de organização (semântico, fonético, morfológico, etc.). As conexões em rede são manifestadas em esquemas: conjunto de palavras que possuem padrões similares de conexões fonológicas ou semânticas das quais emergem generalizações. As regularidades lingüísticas entre os itens são usadas para estruturar o armazenamento e

---

<sup>14</sup> Bybee (2001, p. 30) adota a palavra como unidade de análise e não o morfema (como os estudos tradicionais o fazem), porque, segundo a autora, a palavra tem uma autonomia cognitiva que o morfema não tem. Bybee (2001, p. 30) define palavra como uma unidade de uso que é fonológica e pragmaticamente apropriada em isolado. É interessante notar que a noção de palavra proposta por Bybee (2001, p. 30) inclui, por exemplo, uma seqüência (*chunk*) como, “*Bom dia!*” (significando um cumprimento), que, ortograficamente, é composta por mais de uma palavra. É importante destacar que uma mesma seqüência que, em algumas situações, é analisada como uma palavra (como, “*Bom dia!*”, significando um cumprimento), em outras situações, é analisada como múltiplas palavras. No caso de “*Bom dia!*”, “*bom*” e “*dia*”, ocorrendo separados poderiam ser analisados como duas palavras diferentes e não como uma só palavra, como quando ocorrem juntos na seqüência “*Bom dia!*”. Vale ressaltar que, neste trabalho, consideramos a noção de palavra de Bybee explicitada acima, lembrando que essa noção agrega forma (fonologia) e significado (semântica) e é “motor” central da gramática que tem um léxico forte e dinâmico.

são expressas por meio desses esquemas emergentes. Um exemplo de esquema é apresentado na FIG. 1. Note-se que as palavras deste esquema foram transcritas fonológica e ortograficamente (entre parênteses):

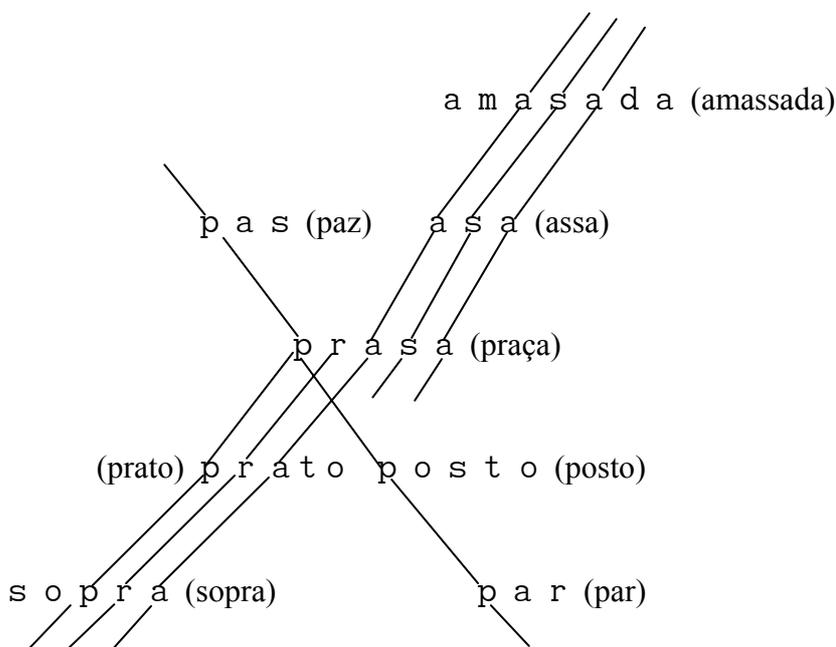


FIGURA 1 – Esquema com conexões lexicais (fonológicas) para [p], [pra] e [asa]

Uma visão nova e importante proposta por Bybee é que categorias, como o som, a sílaba e o traço são graduais (JOHNSON; MULLENIX, 1997) e emergem dos esquemas, portanto, não são dadas a priori. Tais categorias são avaliadas dentro da palavra que é a unidade de análise, não sendo, assim, avaliadas de forma independente, como nas teorias fonológicas tradicionais. Na FIG. 1, isso é ilustrado. A categoria som é uma das que emerge de dentro da palavra, o que pode ser visto nas conexões lexicais que incidem sobre o som [p]. Há ainda, nessa figura, conexões lexicais para a sílaba [pra] e para a seqüência sonora [asa]. As conexões lexicais da FIG. 1 são fonológicas, visto que todas elas se referem apenas ao aspecto sonoro que está presente nas relações

entre as palavras. Essas relações são representadas por meio das linhas transversais em tal figura.

Um outro tipo de conexão lexical proposto por Bybee (2001, p. 22) são as conexões fonológicas e semânticas, das quais emergem relações morfológicas. Estas ocorrem quando há, simultaneamente, similaridades semânticas e fonológicas entre as palavras do esquema (Cf. BYBEE, 2001, p. 22). Um exemplo de conexões fonológicas e semânticas é dado na FIG. 2 a seguir, em que, assim como na FIG. 1, as palavras foram transcritas fonológica e ortograficamente (entre parênteses):

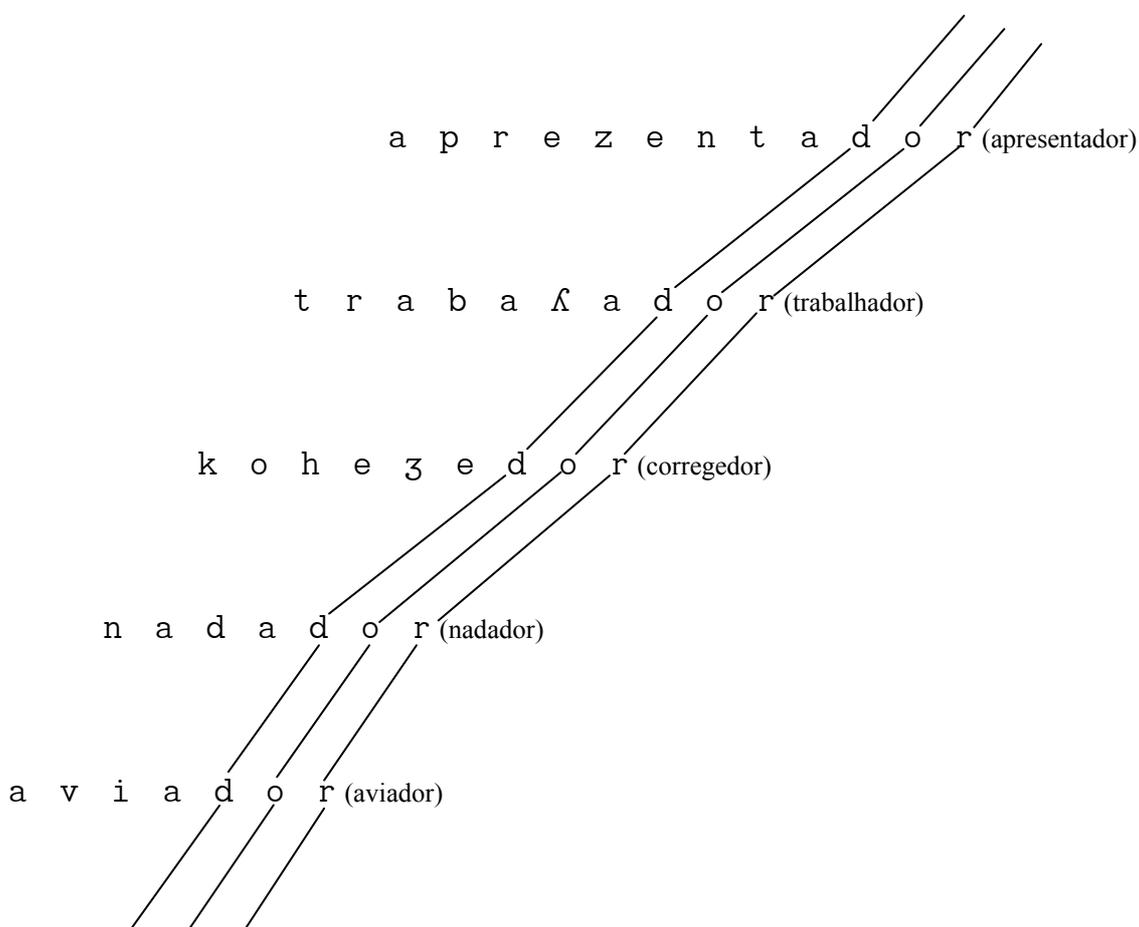


FIGURA 2 – Esquema com conexões lexicais (fonológicas e semânticas – relações morfológicas) para o morfema *-dor*

Como se pode notar pela FIG. 2, o morfema agentivo *-dor* (sufixo nominal)<sup>15</sup> emerge do esquema dessa figura. Pelo fato de emergir um morfema – por meio das similaridades semânticas e fonológicas que as palavras apresentam – é que esse esquema é baseado em relações morfológicas.

Note-se, assim, por meio das FIG. 1 e FIG. 2, que a organização lingüística é feita com base em categorias (como o som, a sílaba e o morfema) que emergem a partir de generalizações inferidas do uso. Por esse motivo, a frequência é crucial na organização da estrutura lingüística, o que será discutido ainda nesta seção.

Em relação à organização do componente fonológico e ao comportamento do indivíduo, é interessante a este trabalho ponderar sobre como funcionam os esquemas. É provável que, como o uso forma os esquemas, estes sejam diferentes no nível individual, de alguma maneira. É claro que usuários de uma mesma língua vão formar esquemas semelhantes. Contudo, a experiência lingüística particular, provavelmente, impedirá que esquemas de falantes de um mesmo grupo sejam idênticos. Falantes têm léxicos e gramáticas diferentes, apesar de compartilharem dos mesmos mecanismos que regem a língua. Um exemplo de um mecanismo que rege a língua é o fato de que a organização e a emergência do léxico e da gramática decorrem das redes lexicais que são formadas por palavras advindas do uso (FIG. 1 e FIG. 2). Certamente que essa idéia de que os indivíduos se diferenciam quanto ao léxico e à gramática pode iluminar questões referentes à variação no comportamento do indivíduo.

---

<sup>15</sup> De acordo com o *Novo Dicionário Eletrônico Aurélio* (HOLANDA, 2004), o morfema sufixo nominal *-dor*, “ocorre em voc. [vocábulos] formados no latim (*abjurador*), ou no vernáculo, a partir de rad. [radical] verbal, com as noções de ‘ofício’, ‘profissão’; ‘que ou aquele que pratica determinada ação (definida pelo rad. verbal)’, ‘agente’, ‘instrumento de ação’: *apreciador*, *carregador*, *trabalhador*; *regador*”. Ainda, segundo Rocha (1998, p. 113), o sufixo *-dor* forma “substantivos-agentivos a partir de verbos: pescador, namorador, franqueador, apelidador [...]”. Vale notar que palavras com esse sufixo se relacionam a um dos três fenômenos em estudo, o Fenômeno 1 (r #), cancelamento de “r” final em nominais (HUBACK, 2003).

A respeito das diferentes gramáticas para diferentes indivíduos, Vihman (1996, p. 5) afirma: “[...] there is, to my knowledge, no evidence that adult speakers of a language share an identical grammar, despite nativist assumptions (see again Mohanan, 1992). On the contrary, individual differences are exhibited by adults [...]”. Ainda Reinecke (2006, p. 91), que analisa os róticos intervocálicos em falantes de Blumenau e Lages, faz a seguinte afirmação:

[...] a língua [é] como um sistema complexo, composta por formas e organizações dessas formas emergentes e representada na gramática individual. As formas, ou os recursos lingüísticos, neste modelo [Fonologia de Uso, de Bybee (2001)], são armazenadas na gramática individual de cada falante. As gramáticas individuais estão permanentemente em processos de adaptação e mudança [...]. A gramática, conforme, entre outros, Wildgen (2005), Selting e Couper-Kuhlen (2000) e Hopper (1987), é ancorada no falante individual.

Outro aspecto importante em que a Fonologia de Uso é diferente das teorias fonológicas anteriores é quanto à consideração do fator frequência (BYBEE, 1995, 2001, 2002). Em Bybee (2001, p. 10), há as seguintes definições para os dois tipos de frequência:

- 1) frequência de ocorrência → é a frequência de ocorrência de uma unidade, geralmente uma palavra, em um texto;
- 2) frequência de tipo → refere-se à frequência de um padrão específico no dicionário, como, por exemplo, um padrão de acento, um afixo ou um encontro consonantal.

Naturalmente que a frequência só pode ser avaliada em uma teoria que leve em conta o uso, por isso ela não havia sido ponderada por teorias fonológicas antecedentes. Bybee (1995) sugere que a frequência de tipo é determinante de produtividade, definida por ela como um padrão que se aplica a novos itens (Cf. BYBEE, 2001, p. 14). A autora afirma

ainda que a interação entre frequência de tipo e frequência de ocorrência é também importante na determinação do grau de produtividade (BYBEE, 1995, 2002).

Essa postura de Bybee é inovadora, na medida em que expande as concepções tradicionais sobre o modo como o léxico e a gramática se estruturam. Primeiro, a autora assume que a palavra é a unidade de armazenamento. Depois, ao atribuir à frequência, seja de tipo ou de ocorrência, um papel fundamental na formação e organização do léxico e da gramática, Bybee aponta que o uso define o léxico, a gramática.

A questão da frequência merece ser ponderada neste trabalho, devido ao fato de ela se relacionar ao comportamento individual: uma palavra como “abóbora” pode ser muito freqüente (em termos de ocorrência) para uma pessoa que gosta muito desse alimento e que o come sempre; mas pode ser pouco freqüente para uma outra pessoa que não gosta dele e que por isso não o come.

Isso implica que, de modo geral, podemos compartilhar de uma gramática comum com a nossa comunidade, porém podemos ter uma organização diferente para essa gramática também, em função de o uso ser sempre particular. A respeito disso, Cristóvão Silva e Abreu Gomes (2004, p. 17), ao discutirem os modelos multirrepresentacionais da Fonologia de Uso, da Teoria de Exemplos e da Lingüística Probabilística, afirmam:

Por termos experiências diferentes como indivíduos, temos também idiosincrasias em cada falante. Por outro lado, por compartilharmos com uma comunidade de fala a construção de nosso conhecimento lingüístico individual, somos capazes de organizá-lo num sistema plástico e dinâmico.

Ainda, em relação à questão da frequência associada ao comportamento do indivíduo, Oliveira Guimarães (2004, p. 45) aponta:

Essa teoria [Fonologia de Uso] possui alguns aspectos ainda não bem definidos, como por exemplo, a relação da frequência com o indivíduo. Afinal, cada indivíduo tem uma bagagem linguística diferente, dependendo do seu meio, sua profissão etc. Portanto, é esperado que a frequência de ocorrência dos itens lexicais varie de indivíduo para indivíduo. Apesar disso (e de outros problemas os quais não caberia serem tratados aqui), essa teoria traz contribuições interessantes ao considerar o uso como um fator importante na propagação da variação e mudança sonora no léxico. Certamente, o desenvolvimento da linguística de corpora e o estudo da variação e mudança sonora sob essa perspectiva poderão contribuir para o aprimoramento dessa teoria.

Outra idéia que vale ser destacada de Bybee (2001) é a afirmação desta autora de que a mudança sonora é fonética e lexicalmente gradual. Isso pode ser confrontado com o Modelo Neogramático, segundo o qual a mudança sonora é foneticamente gradual e lexicalmente abrupta (OSTHOFF; BRUGMAN, 1878). E pode ainda ser confrontado com o Modelo de Difusão Lexical (CHEN; WANG, 1975), segundo o qual a mudança sonora é foneticamente abrupta e lexicalmente gradual. É interessante observar que Bybee (2002) defende que há padrões de difusão lexical nos casos de mudança sonora. Assim, podemos notar que, nesse aspecto de padrões de difusão lexical para casos de mudança sonora, a Fonologia de Uso, de Bybee, se relaciona com o Modelo de Difusão Lexical (CHEN; WANG, 1975) e o complementa ao sugerir a gradualidade fonética.

Ainda a respeito das representações múltiplas, ao contrário das teorias tradicionais – como se apontou nesta seção –, o modelo de Bybee é maximamente redundante no léxico. A representação é redundante, feita por meio de exemplares (balizados pela frequência de ocorrência), com detalhamento fonético e generalizações que emergem dos tipos: vogal, consoante, *onset*, sílaba, etc. Desse modo, em sendo as representações múltiplas, o componente linguístico convive com a variação no léxico. Mais uma vez, nota-se que a Fonologia de Uso permite que se proponha o comportamento do indivíduo

como unidade de análise. Somente uma teoria que reconhece a variação no léxico (variação esta que é decorrente da existência de representações múltiplas), é capaz de lidar com as variações inter e intra-individuais.

É bom lembrar ainda que, dentro de teorias fonológicas tradicionais, a representação fonológica é vista como algo abstrato, independente da experiência. Na Fonologia de Uso, diferentemente, a representação fonológica se organiza a partir do mapeamento da experiência. As representações mentais das categorias e padrões fonológicos são gradualmente construídas por meio da experiência de fala. Então, generalizações emergem e formam entidades de caráter abstrato. Mesmo com esse caráter abstrato, ainda assim, nota-se que, na teoria de Bybee, a representação fonológica é consolidada pelo uso. Por ser esta uma pesquisa em fonologia, é importante para o presente trabalho abordar as representações fonológicas que vão organizar o componente fonológico. Uma única palavra pode ter representações fonológicas diferentes para indivíduos diferentes ou mesmo para um único indivíduo. Isso caracteriza a existência de representações múltiplas na organização do componente fonológico.

Por toda a explanação realizada nesta seção, é possível perceber o quanto a Fonologia de Uso, proposta por Bybee (2001), é fundamental ao presente estudo. Esse modelo multirrepresentacional oferece o instrumental teórico necessário para a análise do comportamento do indivíduo (variações inter e intra-individuais) em diversos fenômenos fonológicos. Isso é possibilitado por aspectos dessa teoria abordados nesta seção, como: a experiência afetar a organização lingüística; a palavra ser a unidade de análise; a frequência (de ocorrência e de tipo) ser crucial na organização do componente fonológico; as representações incorporarem a redundância com o detalhe fonético, etc.

Vale destacar ainda que, ao postular a Fonologia de Uso, Bybee se utiliza da Teoria de Exemplos, proposta por Johnson (1997) e complementada por Pierrehumbert (2001b), a fim de caracterizar os esquemas representacionais em seu modelo (FIG. 1 e FIG. 2). Assim, na próxima seção, tratamos da Teoria de Exemplos que se conecta com a Fonologia de Uso.

### 3.3 Teoria de Exemplos

Passemos a abordar a Teoria de Exemplos na proposta de Pierrehumbert (2001b). A Teoria de Exemplos foi introduzida na área da psicologia como um modelo de representação e categorização (PIERREHUMBERT, 2001b, p. 139). Esse modelo teórico foi estendido para a área de estudos dos sons da fala, com Goldinger (1996) e Johnson (1997), sendo que Pierrehumbert (2001b), trabalho aqui enfatizado, propõe importantes avanços a tal modelo.

Segundo a Teoria de Exemplos (PIERREHUMBERT, 2001b), cada categoria – como, por exemplo, o som, a sílaba ou o morfema – é representada na memória por uma “nuvem” de ocorrências da categoria. Na FIG. 3, veja-se o exemplo de uma nuvem de ocorrências/ exemplos para o morfema agentivo *-dor* (Cf. nota 15 na seção anterior). Vale lembrar que tal morfema, o qual é relacionado ao Fenômeno 1 (r #) em análise (QUADRO 1, capítulo 1), foi utilizado para demonstrar a representação de um esquema lexical na FIG. 2 da seção anterior.

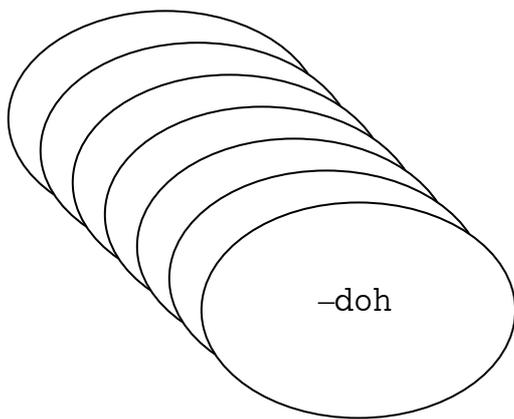


FIGURA 3 – Nuvem de exemplares para o sufixo *-dor*

Na nuvem de exemplares da FIG. 3, a categoria é o morfema *-dor*. A FIG. 4, a seguir, apresenta uma nuvem de múltiplos exemplares para esse morfema:

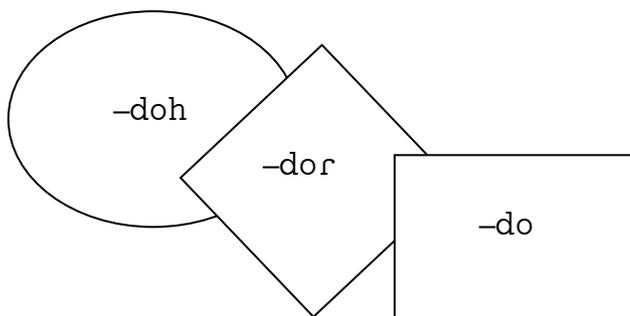


FIGURA 4 – Nuvem de múltiplos exemplares para o sufixo *-dor*

A FIG. 4 apresenta uma nuvem de exemplares com múltiplas representações para o sufixo *-dor*: [doh], [dor] e [do]. Estas três formas são pronúncias variantes possíveis para tal morfema no português brasileiro.

Segundo Pierrehumbert (2001b), na nuvem, cada ocorrência de uma categoria é armazenada como um exemplar. Desse modo, categorias mais frequentes vão ser

representadas por muitos exemplares que serão mais ativados, enquanto que categorias menos freqüentes vão ser representadas por menos exemplares que serão menos ativados. Veja-se o exemplo da FIG. 5 que apresenta uma nuvem de exemplares para o sufixo *-dor*, contudo, com as formas variantes [doh] e [do] em competição:

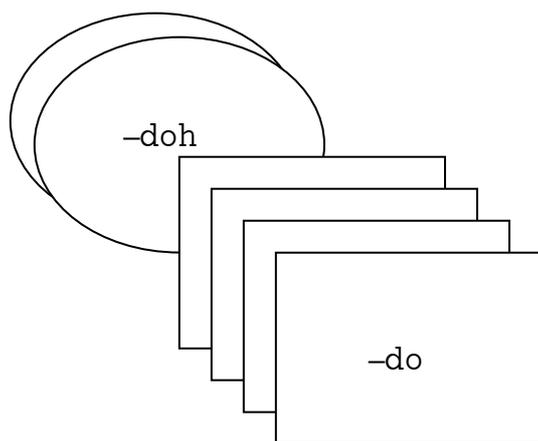


FIGURA 5 – Nuvem de exemplares para o sufixo *-dor*, com [-doh] e [-do] em competição (Baseado em CRISTÓFARO SILVA, 2002)

Como se pode observar pela análise da FIG. 5, as formas [-doh] e [-do] se encontram em competição, de modo que a forma *-do* apresenta mais ocorrências, sendo, portanto, mais freqüente e mais ativada do que a forma [-doh]. As categorias não são, como se pode notar, pré-estabelecidas. Elas são generalizações dinâmicas de dados lingüísticos. O caráter robusto das categorias depende da freqüência de ocorrência. Categorias mais freqüentes são mais robustas. Já categorias menos freqüentes são menos robustas (Cf. PIERREHUMBERT, 2003a). No caso da nuvem de exemplares da FIG. 5, a categoria mais robusta é a da forma [-do] que é mais freqüente, sendo a forma [-doh], ao contrário, menos robusta por ser menos freqüente.

A respeito da categorização de exemplares, Pierrehumbert (2001b, p. 140) afirma que, quando uma nova ocorrência é encontrada, ela é classificada, nessa teoria, de acordo com a sua similaridade com exemplares já armazenados. Neste ponto, o Modelo de Exemplares se aproxima da Fonologia de Uso por estocar e organizar as ocorrências por similaridades (Cf. seção 3.2), neste caso, fonéticas.

Um outro princípio do Modelo de Exemplares é o de que variantes que ocorrem mais freqüentemente constituem o protótipo para a categoria. Como se pode notar, a Teoria de Exemplares mantém ligação com a Fonologia de Uso, em especial na questão da freqüência, considerada pelo presente trabalho. Essa formação de exemplares, sendo dependente da freqüência de ocorrência, pode ter relação com o comportamento do indivíduo. É provável que diferentes indivíduos de um mesmo grupo social apresentem exemplares e organização de exemplares diferentes.

Ainda, o Modelo de Exemplares, assim como a Fonologia de Uso (Cf. BYBEE, 2001, p. 51, 52), supõe que a memória humana tem uma capacidade enorme. Isso está em consonância com as descobertas atuais da ciência sobre o cérebro humano (Cf. SCHACTER; CHURCH, 1992; PALMERI; GOLDINGER; PISONI, 1993; JOHNSON, 1997; SZCZENIAK, 2004). A esse respeito, Pierrehumbert (2001b, p. 139) comenta: “The mind’s capacity for longterm memories of individual examples is in fact astonishingly large, as experiments reviewed in Johnson (1997) indicate”. Também Cristóvão Silva e Abreu Gomes (2004, p. 17), tratando da Fonologia de uso, do Modelo de Exemplares e da Lingüística Probabilística, afirmam: “Quaisquer que sejam os limites da memória, se as ocorrências não fossem armazenadas, nenhum protótipo ou exemplar poderia ser formado, uma vez que a abstração é determinada pela comparação de múltiplas unidades individuais”.

A consideração de que a memória tem uma capacidade enorme é essencial à investigação do comportamento do indivíduo, foco deste trabalho. Como propomos que a organização do componente fonológico de um indivíduo é baseada em representações múltiplas, é necessário, para isso, uma grande capacidade de memória.

Pierrehumbert (2001b, p. 142) aponta:

Exemplar theory provides us with a way to formalize the detailed phonetic knowledge that native speakers have about the categories of their language. Since exemplar theory stores directly the distribution of phonetic parameter values associated with each label, it provides us with a picture of the “implicit phonetic knowledge of the speaker”.

É justamente pelo fato de o Modelo de Exemplares oferecer a possibilidade de se formalizar o conhecimento fonético detalhado, como afirma Pierrehumbert, que ele é importante a este estudo. Esse conhecimento pressupõe múltiplas representações (FIG. 4 e FIG. 5) que, como se disse logo anteriormente, é uma noção básica para a investigação do comportamento do indivíduo em relação a fenômenos de variação e mudança sonora e, também, para o mapeamento da organização fonológica.

Ainda, essa idéia do conhecimento fonético detalhado se liga à outra proposta do Modelo de Exemplares: a consideração de que o segmento sonoro é gradiente. A gradualidade fonética é, inclusive, avaliada acusticamente. De acordo com Pierrehumbert (2003a), os fonemas são distribuições probabilísticas dentro de um espaço fonético contínuo. Assim, itens que têm uma distribuição central seriam bons exemplos de um fonema particular, enquanto que itens periféricos na distribuição, seriam marginais como membros. Essa noção de que o segmento sonoro é gradiente se liga à idéia das representações múltiplas, as quais são organizadas por meio de semelhanças fonéticas entre os exemplares. Sendo assim, tal noção é importante para

este estudo. A variação incorporada nas representações (FIG. 4 e FIG. 5), o que a gradiência sonora supõe, permite que se investigue o comportamento do indivíduo segundo variações inter e intra-individuais.

É bom lembrar que a formação de exemplares é dinâmica, já que o léxico é o tempo inteiro atualizado (BYBEE, 2002). Considerando isso, podemos dizer que a Teoria de Exemplares, juntamente com a Fonologia de Uso e a Lingüística Probabilística – a qual é apresentada na próxima seção – se apresentam como a base teórica mais adequada para tratar das variações inter e, especialmente, intra-individuais. Isso porque tais teorias prevêm variações e mudanças contínuas no léxico individual.

Sabemos que há limitações metodológicas na avaliação dessas variações, já que não é possível acompanhar cada indivíduo ao longo de sua existência. Até mesmo Pierrehumbert (2001b, p. 139) afirma: “[...] the volume of speech which a person processes in a lifetime is so great that we would not wish to assume individual memories of every use of every word.” Ainda assim, esta pesquisa pode iluminar essa questão, ao avaliar detidamente a produção lingüística de indivíduos específicos.

Oliveira Guimarães (2004, p. 40) é um trabalho que vale ser citado nesta seção, por fazer uma interessante comparação entre a proposta teórica tradicional e o Modelo de Exemplares, o que pode ser visto no QUADRO 2:

**QUADRO 2:** Comparação entre a proposta tradicional e o Modelo de Exemplos

<b>Proposta tradicional</b>	<b>Modelo de Exemplos</b>
Representação mental minimalista	Representação mental detalhada
Separação entre fonética e fonologia	Inter-relação da fonética e da fonologia
Visão da fonologia como uma gramática formal, com a utilização de variáveis abstratas	Consideração de que a fonologia da língua envolve a distribuição probabilística de variáveis
Efeitos da frequência refletidos na produção em curso e não armazenados na memória de longa duração	Efeitos da frequência armazenados na memória de longa duração
Julgamento fonotático categórico: uma seqüência ou é considerada bem formada ou é impossível de ocorrer na língua	Efeitos gradientes nos julgamentos fonotáticos
Léxico separado da gramática fonológica	Palavra como <i>locus</i> da categorização

Fonte: Oliveira Guimarães (2004, p. 40).

Na primeira coluna do QUADRO 2, são apresentadas características da proposta teórica tradicional. Já a segunda coluna apresenta características do Modelo de Exemplos. Por meio da análise do QUADRO 2, nota-se, mais uma vez, que o Modelo de Exemplos mantém estreita relação com a Fonologia de Uso. Na verdade, todas as características apresentadas para o Modelo de Exemplos podem ser atribuídas aos outros dois modelos multirrepresentacionais em análise: Fonologia de Uso e Lingüística Probabilística. Esta última teoria é avaliada na próxima seção.

### 3.4 Lingüística Probabilística

O livro de Bod, Hay e Jannedy (2003) reúne textos de diferentes autores que trabalham com a Lingüística Probabilística. Nesse livro, são realizadas abordagens probabilísticas para diferentes áreas da lingüística: Psicolingüística, Sociolingüística, Mudança Lingüística, Fonologia, Morfologia, Sintaxe e Semântica.

A Lingüística Probabilística nasce da idéia de que a “language is probabilistic” (BOD; HAY; JANNEDY, 2003, p. vii). Como afirmam Bod, Hay e Jannedy (2003, p. 7), “Probabilities permeate the linguistic system”. No caso da fonologia, Bod, Hay e Jannedy (2003, p. vii) apontam: “In phonology [...] probabilities play a role in acceptability judgments and alternations”.

Essa noção de que a probabilidade permeia o sistema lingüístico é fundamental para a abordagem do comportamento do indivíduo, proposta neste estudo. É praticamente impossível contabilizar, uma a uma, todas as ocorrências armazenadas no léxico de um falante, ao longo de sua vida. Contudo, a Lingüística Probabilística oferece o instrumental que viabiliza o tratamento estatístico dessas ocorrências e, com isso, permite que se analisem questões – como a frequência, por exemplo –, as quais não seriam passíveis de análise sem o auxílio da probabilidade. Bod, Hay e Jannedy (2003, p. 1) afirmam: “Probabilistic linguistics increases the range of data for which a theory can account, and for which it must be accountable”. Ainda, como “[...] entities at all levels of representation in phonetics and phonology display statistical variation” (PIERREHUMBERT, 2003b, p. 226), é imprescindível, a este trabalho, lançar mão da Lingüística Probabilística na investigação do comportamento do indivíduo.

A Linguística Probabilística questiona a noção de categorias fixas e propõe que: “Categories are central to linguistic theory, but membership in these categories need not be categorical. Probabilistic linguistics conceptualizes categories as distributions. Membership in categories is gradient” (BOD; HAY; JANNEDY, 2003, p. 4). Com isso, um ponto fundamental da Linguística Probabilística emerge: a noção de gradiência em um contínuo:

The probabilistic approach to the study of language may seem opposed to the categorical approach, which has dominated linguistics for over 40 years. [...] while categorical approaches focus on the endpoints of distributions of linguistic phenomena, probabilistic approaches focus on the gradient middle ground. (BOD; HAY; JANNEDY, 2003, p. vii).

Ainda sobre a gradiência, que é muito importante dentro da Fonologia de Uso e da Teoria de exemplares também, os autores mencionam:

Language displays all the hallmarks of a probabilistic system. Categories and well-formedness are gradient, and frequency effects are everywhere. We believe all evidence points to a probabilistic language faculty. Knowledge of language should be understood not as a minimal set of categorical rules or constraints, but as a (possibly redundant) set of gradient rules, which may be characterized by a statistical distribution. (BOD; HAY; JANNEDY, 2003, p. 10).

Como se apontou na seção anterior, a idéia da gradiência auxilia na investigação do comportamento do indivíduo, uma vez que se relaciona com a noção de representações múltiplas. Tratando da probabilidade na variação, até mesmo no nível individual, os autores afirmam:

Within individuals, production patterns differ on the basis of stylistic factors such as addressee, context, and topic, and this stylistic variation to a large degree echoes the variation present across members of society. [...] And individuals differ not only in the specific variants they use in different contexts, but also in the

frequency with which they use them. Knowledge of variation must involve knowledge of frequencies. (BOD; HAY; JANNEDY, 2003, p. 2, 3).

Nessa última citação, Bod, Hay e Jannedy (2003) lançam uma questão interessante a ser considerada pelo presente trabalho: indivíduos se diferem não somente quanto a variantes específicas que usam em contextos diferentes, mas também quanto à frequência com que usam essas variantes. Certamente que podem ser observadas variações no comportamento do indivíduo quanto à frequência de uso de variantes. Indivíduos de um mesmo grupo social podem fazer uso de uma mesma forma, mas com frequências diferentes. Conseqüentemente, isso influencia na organização do componente fonológico de cada indivíduo de um único grupo, uma vez que uma mesma palavra pode apresentar muitos exemplares para um indivíduo e poucos exemplares para outro. Sendo assim, mais uma vez, constatamos que a frequência é muito importante ao estudo da variação sonora, como também apontam Bod, Hay e Jannedy (2003) na última citação.

Ainda em relação à frequência, questão importantíssima na Lingüística Probabilística – e também na Fonologia de Uso e no Modelo de Exemplares, como se observou nas seções 3.2 e 3.3 –, Bod, Hay e Jannedy (2003, p. 3) afirmam:

One striking clue to the importance of probabilities in language comes from the wealth of frequency effects that pervade language representation, processing, and language change.

[...] frequency permeates language.

[...] We can best model many of these effects by making explicit the link between frequency and probability. Probability theory provides well-articulated methods for modeling frequency [...].

Os autores chegam a apontar que a frequência é uma espécie de pré-requisito para a existência da Lingüística Probabilística:

Just as the complete absence of frequency effects would challenge the foundations of probabilistic linguistics, so their overwhelming presence adds weight to the claim that the language faculty is inherently probabilistic.

[...] Frequency effects provide one type of evidence for a probabilistic linguistics (BOD; HAY; JANNEDY, 2003, p. 3, 4).

Por toda a discussão realizada nesta seção acerca da Lingüística Probabilística, percebe-se que esta teoria, a Fonologia de Uso e o Modelo de Exemplares se apresentam como a base teórica mais adequada ao estudo do comportamento do indivíduo em diversos fenômenos fonológicos. Como apontado, a Lingüística Probabilística oferece o instrumental para se analisar a língua estatisticamente, o que é mais apropriado, devido ao caráter probabilístico da língua. No caso deste estudo sobre o comportamento do indivíduo, a Lingüística Probabilística permite que se investiguem as múltiplas representações do componente fonológico, decorrentes das variações inter e intra-individuais. Essencialmente, a Lingüística Probabilística é importante ao presente trabalho, na medida em que ela permite um tratamento probabilístico às representações múltiplas postuladas pela Fonologia de Uso e pela Teoria de Exemplares.

### **3.5 Sumário**

Este capítulo apresentou e discutiu a base teórica que fundamenta a presente pesquisa: Fonologia de Uso (BYBEE, 2001), Teoria de Exemplares (PIERREHUMBERT, 2001b) e Lingüística Probabilística (BOD; HAY; JANNEDY, 2003). Tais modelos multirrepresentacionais – que consideram que as representações lingüísticas são complexas e que o mapeamento entre as formas abstratas e as formas atestadas nas

línguas naturais é simples – são importantes a este estudo, por permitirem que o comportamento do indivíduo seja proposto como unidade de análise.

A respeito da Fonologia de Uso, como destacamos, esse modelo oferece um instrumental teórico muito relevante para a análise das variações inter e intra-individuais quanto a diversos fenômenos fonológicos. Isso é o que vimos ao discutir aspectos dessa teoria, como: a experiência afetar a organização lingüística; a palavra ser a unidade de análise; as frequências de ocorrência e de tipo serem fundamentais na organização do léxico e da gramática; as representações apresentarem redundância com o detalhe fonético, etc.

Apontamos que a Teoria de Exemplos, que é incorporada pela Fonologia de Uso, também é importante ao presente trabalho. Isso ocorre, em especial, pelo fato de o Modelo de Exemplos considerar a gradiência fonética que se liga à noção de representações múltiplas, a qual é elucidativa na abordagem da variabilidade individual.

Sobre a Lingüística probabilística, como ressaltamos, por propor uma análise probabilística da língua, essa teoria permite que se investiguem as múltiplas representações do componente fonológico, decorrentes das variações inter e intra-individuais. Tal teoria possibilita um tratamento estatístico mais refinado para as representações múltiplas postuladas pela Fonologia de Uso e pela Teoria de Exemplos.

Passemos agora ao próximo capítulo, o qual aborda a metodologia empregada neste estudo.

## **CAPÍTULO 4**

### **METODOLOGIA**

#### **4.1 Introdução**

Este capítulo apresenta a metodologia empregada nesta pesquisa. A seção 4.2 aponta o tipo de método adotado. A seção 4.3 apresenta a amostra trabalhada. A seção 4.4 trata do material utilizado, de todo o aparato e estímulos. A seção 4.5 apresenta os procedimentos de coleta e análise dos dados dos estudos transversal e longitudinal aqui realizados. Por último, a seção 4.6 sumariza este capítulo.

#### **4.2 Tipo de método adotado**

Deve-se destacar que não há uma metodologia específica para se avaliar a variabilidade no comportamento do indivíduo até o atual momento. O presente trabalho sugere procedimentos metodológicos – descritos no decorrer deste capítulo – que se pretende que sejam mais adequados ao tratamento das variações individuais.

Dessa forma, a metodologia aqui proposta deve possibilitar a análise do comportamento lingüístico do indivíduo, considerando que esse comportamento é baseado em representações múltiplas (BYBEE, 2001, 2002). Neste caso, não podemos ter uma

hipótese exata sobre o comportamento de cada indivíduo em relação à realização dos fenômenos aqui analisados, os quais serão discutidos na seção 4.4.1. Assim, por exemplo, quanto ao cancelamento de “r” final em nominais (ex.: cantor → cantô), não podemos afirmar se um determinado indivíduo irá ou não realizar tal fenômeno ou se, realizando-o, irá cometer mais ou menos cancelamentos.

Ao considerarmos que o comportamento do indivíduo é baseado em representações múltiplas, pautado na experiência, no uso, é dinâmico e atualizado de forma *on line* (BYBEE, 2001), supomos que ele deve ser, de certa forma, imprevisível. Assim, novamente, não podemos formular uma hipótese forte a respeito do comportamento do indivíduo. Por esse motivo, adotamos como método o Estudo Descritivo Exploratório. Neste tipo de método, apenas se descrevem os resultados encontrados sem se buscar a confirmação de uma hipótese pré-estabelecida.

Esta pesquisa conta com uma abordagem de tempo aparente, já que é realizado um estudo transversal com os 12 indivíduos aqui selecionados. O capítulo 5 aborda tal estudo.

Ainda, o presente trabalho conta com uma abordagem de tempo real, já que apresenta três análises de caráter longitudinal, as quais podem ser conferidas no capítulo 6. Na primeira análise, avalia-se se houve mudança estatisticamente significativa, para os 12 indivíduos em estudo, do corpus de origem de cada um deles (corpora 1, 2 e 3) para o corpus atual. Essa análise comparativa tem, assim, um caráter longitudinal, visto que observa a produção lingüística dos mesmos indivíduos em tempos diferentes: em pesquisas anteriores e na atual. A segunda análise incide sobre o estudo longitudinal de diversas interações aqui realizado; para sermos mais exatos, de 6 interações ou coletas

mensais de dados, com 2 dos 12 indivíduos do estudo transversal. Na terceira análise de caráter longitudinal que fazemos, observamos esses 2 indivíduos do estudo longitudinal de 6 coletas, comparando a produção lingüística deles no estudo transversal e no estudo longitudinal realizados neste trabalho.

Sobre o estudo da mudança em tempo real, Paiva e Duarte (2003, p. 17) fazem a seguinte afirmação:

O estudo da mudança em tempo real (de curta ou longa duração) permite recobrir aspectos que não podem ser detectados pelo estudo em tempo aparente, distinguindo mudanças que se produzem de forma gradual em toda a comunidade lingüística daquelas que podem caracterizar a trajetória de comportamento lingüístico do indivíduo ao longo da sua vida. (PAIVA; DUARTE, 2003, p. 17).

Por essa afirmação de Paiva e Duarte (2003), é possível notar que um estudo em tempo real é muito importante a este trabalho, por nos possibilitar identificar possíveis variações no comportamento do indivíduo. Podemos dizer ainda que, mais do que estarmos fazendo um estudo em tempo real, estamos realizando um estudo de painel: re-entrevistas com os mesmos indivíduos sobre um período de tempo (Cf. BAILEY, 2002, p. 329).

Segundo Bailey (2002, p. 329), dados desse tipo – obtidos em tempo real, em uma pesquisa do tipo estudo de painel – são cruciais para se explorar estabilidade e mudança em vernáculos individuais. Vale destacar, também, o comentário de Paiva e Duarte (2003, p. 17) sobre esse tipo de pesquisa: “O estudo de painel, através da comparação de amostras de fala dos mesmos falantes em diferentes pontos do tempo, permite captar mudanças ou estabilidade no comportamento lingüístico do indivíduo [...]”. É interessante retomar a pesquisa do Programa de Estudos sobre o Uso da Língua – PEUL, a qual foi abordada no capítulo 2. Como foi visto, nos dados desse estudo, os

quais provêm de duas entrevistas coletadas em dois tempos diferentes com os mesmos informantes (estudo em tempo real), foram constatadas variações inter e intra-individuais.

A proposta de se utilizar aqui o estudo longitudinal na análise da variabilidade individual é também motivada pelo fato de que trabalhos em aquisição da fonologia adotam tal estudo para observar variações individuais decorrentes do próprio desenvolvimento da linguagem infantil<sup>16</sup>. É sabido que toda criança, durante o processo de aquisição de uma língua, apresenta variações na sua fala (Cf. VIHMAN, 1996; DEMUTH, 1997; HAZAN; BARRETT, 2000), devido ao fato de que ela está se desenvolvendo lingüisticamente, de que ela ainda não adquiriu a língua.

Assim, observa-se que, nos trabalhos de aquisição, a adoção do estudo longitudinal como procedimento metodológico permite que se avaliem melhor as variações no comportamento do indivíduo (OLIVEIRA GUIMARÃES, em andamento). Contudo, é preciso dizer que esses trabalhos em aquisição não se interessam pela questão da variabilidade individual. Antes, neles, tal variação é analisada em função somente do desenvolvimento na aquisição da linguagem infantil.

Ainda, é preciso dizer que o presente estudo adota aspectos da Sociolingüística (LABOV, 1966), já que os fenômenos em análise apresentam-se como casos de variação. Dentre esses aspectos que utilizamos, podemos citar, por exemplo, a adoção da entrevista de fala espontânea e das tarefas de leitura e de nomeação por figura nesta pesquisa.

---

<sup>16</sup> Para uma maior discussão a respeito da adoção do estudo longitudinal em trabalhos sobre a fala infantil e em trabalhos sobre a fala adulta, consultar Fontes Martins; Oliveira Guimarães (em andamento).

### 4.3 Amostra

Os sujeitos deste trabalho foram selecionados dos corpora 1, 2 e 3<sup>17</sup> apresentados no QUADRO 1 (Cf. capítulo 1). Na seleção de tais sujeitos, consideramos como principal critério a questão da liderança em relação à mudança sonora (LABOV, 2001, v. 2). Como foi apontado no capítulo 2, a classificação de um falante como líder ou não-líder seguiu a proposta de Labov (2001, v. 2) e se baseia nos índices estatísticos, obtidos para esse falante, de ocorrência de uma variável inovadora em caso de variação sonora. Um líder apresenta maiores índices da realização de uma variável inovadora do que um falante não-líder. Assim, foram escolhidos, em cada um dos três corpora aqui considerados, os 2 falantes que mais realizaram um fenômeno em questão – indivíduos líderes – e os 2 falantes que menos realizaram esse fenômeno – indivíduos não-líderes. No total, foram selecionados 12 informantes: 6 líderes e 6 não-líderes.

Como se pode perceber, a orientação central, nesse procedimento metodológico de escolha dos falantes, foi agrupar separadamente indivíduos líderes e não-líderes da mudança sonora (LABOV, 2001, v. 2). Assim, na seleção desses 12 informantes, não houve, primariamente, preocupação com fatores, como classe social, sexo, idade e escolaridade do falante.

O QUADRO 3 a seguir apresenta características dos 12 informantes desta pesquisa, segundo fatores que são apontados na primeira linha de tal quadro. A codificação desses fatores é esclarecida no QUADRO 4.

---

<sup>17</sup> Os corpora 1, 2 e 3 serão discutidos na seção 4.4.1, com maiores detalhamentos.

**QUADRO 3: Informantes**

<b>Informante</b>	<b>Corpora</b>	<b>Liderança</b>	<b>Sexo</b>	<b>Faixa etária</b>	<b>Escolaridade</b>	<b>Classe social</b>
1.	1	L	M	J	X	B
2.	1	L	M	I	X	B
3.	1	N	F	V	Z	A
4.	1	N	M	V	Z	A
5.	2	L	M	J	Z	A
6.	2	L	F	J	Z	A
7.	2	N	M	V	Z	A
8.	2	N	F	V	Z	A
9.	3	L	M	I	Z	A
10.	3	L	M	V	X	B
11.	3	N	F	V	Z	A
12.	3	N	F	I	Z	A

**QUADRO 4: Codificação referente ao QUADRO 3**

	<b>Categorias</b>	<b>Códigos</b>
CORPORA	Corpus 1	1
	Corpus 2	2
	Corpus 3	3
LIDERANÇA	Líder	L
	Não-líder	N
SEXO	Masculino	M
	Feminino	F
FAIXA ETÁRIA	Jovem: 18-30 anos	J
	Intermediário: 31-45 anos	I
	Velho: 46-60 anos	V
ESCOLARIDADE	Fundamental	X
	Médio	Y
	Superior	Z
CLASSE SOCIAL	Baixa	B
	Média	A

O QUADRO 3 discrimina, na primeira coluna, os 12 sujeitos desta pesquisa. Da segunda a sétima e última coluna, tal quadro apresenta, para cada informante, uma classificação baseada nos seguintes critérios: a) corpora – 2<sup>a</sup> coluna; b) liderança – 3<sup>a</sup> coluna; c) sexo – 4<sup>a</sup> coluna; d) faixa etária – 5<sup>a</sup> coluna; e) escolaridade – 6<sup>a</sup> coluna; f) classe social – 7<sup>a</sup> coluna.

Somente as segunda e terceira colunas apresentam uma distribuição simétrica para os fatores apresentados na primeira linha do QUADRO 3. Assim, pela análise da segunda coluna deste quadro, podemos notar que há 4 informantes para cada um dos 3 corpora. Na terceira coluna, percebemos que há exatamente 2 indivíduos líderes e 2 indivíduos não-líderes para cada corpora.

A partir da terceira coluna do QUADRO 3, a distribuição é assimétrica: por exemplo, a quarta coluna mostra que há 3 indivíduos do sexo masculino e 1 do sexo feminino no Corpus 1. Isso indica a assimetria citada. Geralmente, estudos que adotam a metodologia sociolinguística fielmente, em todos os aspectos, selecionam os informantes de forma simétrica em relação aos fatores apontados na primeira linha do QUADRO 3 (por exemplo, se são selecionados 50 informantes, quanto ao sexo, deve haver 25 homens e 25 mulheres), apresentando ainda, como recomendado, um mínimo de 5 informantes por célula social (LABOV, 1975).

Vale notar que, pela análise da quarta coluna do QUADRO 3, o número de homens (7) é superior ao de mulheres (5). Quanto à faixa etária (quinta coluna do QUADRO 3), é possível perceber que: há 3 informantes na faixa jovem (de 15-30 anos de idade), 3 informantes na faixa intermediário (de 31-45 anos de idade) e 6 informantes na faixa velho (de 46-60 anos de idade).

Em relação à escolaridade (sexta coluna do QUADRO 3), pode-se notar que há 3 informantes com nível fundamental e 9 informantes com nível superior. Por último, quanto à classe social (sétima coluna do QUADRO 3), percebe-se que há 3 informantes de classe baixa e 9 informantes de classe média. Convém destacar que, como se pode

observar nas sexta e sétima colunas do QUADRO 3, os 3 informantes que têm ensino fundamental são exatamente os 3 informantes de classe baixa, assim como os 9 informantes de nível superior de escolaridade são os 9 informantes de classe média. Isso demonstra uma tendência que é apontada em estudos sobre mudança e variação lingüística: em geral, a escolaridade e a classe social de um indivíduo estão ligadas (Cf. BAGNO, 2002). Quanto menor a classe social de uma pessoa, menor a sua escolaridade e vice-versa.

Para sumarizar as informações discutidas anteriormente a respeito dos participantes desta pesquisa, apresentamos o QUADRO 5:

**QUADRO 5: Líderes e não-líderes**

	Líderes			Não-líderes		
Sexo	1 mulher	5 homens		4 mulheres	2 homens	
Faixa Etária	3 jovens	2 intermediários	1 velho	0 jovens	1 intermediário	5 velhos
Classe Social	3 baixa	3 média		0 baixa	6 média	
Escolaridade	3 fundamental	0 média	3 superior	0 fundamental	0 média	6 superior

O QUADRO 5 permite uma visualização melhor da diferença entre líderes (2ª coluna) e não-líderes (3ª coluna), quanto aos fatores sexo, faixa etária, classe social e escolaridade (1ª coluna). É relevante fazer aqui essa diferenciação, visto que a escolha dos participantes dessa pesquisa teve como critério a questão da liderança, o que foi destacado ainda nesta seção. As células destacadas em cinza, em tal quadro, não apresentaram nenhum indivíduo.

Pela análise do QUADRO 5, podemos fazer algumas observações sobre os participantes líderes e não-líderes desta pesquisa:

- de certa maneira, contrariando Labov (2001, v. 2), os líderes são, na maioria, homens (5 homens e apenas 1 mulher). Já os não-líderes são, na maioria, mulheres (4 mulheres e somente 2 homens);
- a maior parte dos líderes é jovem (3 indivíduos). Não há um único jovem não-líder; a grande maioria dos não-líderes é composta de velhos (5 indivíduos);
- metade dos indivíduos líderes é de classe baixa, e a outra metade é de classe média. Já quanto aos não-líderes, todos são de classe média;
- quanto à escolaridade dos indivíduos líderes, há metade com nível fundamental (note-se que todos eles são de classe baixa) e metade com nível superior (os indivíduos com nível superior são exatamente os de classe média). Todos os indivíduos não-líderes têm nível superior (todos com nível superior são de classe média).

Resumindo, podemos dizer que os líderes são, na sua maioria, homens, jovens, sendo que uma metade deles é proveniente de classe baixa, enquanto que a outra metade é proveniente de classe média; assim como uma metade apresenta nível fundamental de escolaridade, enquanto a outra metade apresenta nível superior de escolaridade. Já os indivíduos não-líderes são, na sua maioria, mulheres, velhos, todos de classe média e de nível de escolaridade superior.

No QUADRO 6, apresentamos a codificação que é adotada para os 12 participantes neste trabalho. Essa codificação será importante especialmente nos próximos dois capítulos de análise dos dados, pois será utilizada muitas vezes, em tal capítulo.

**QUADRO 6: Codificação dos participantes**

<b>Código do participante</b>	<b>Liderança</b>	<b>Corpus de proveniência</b>	<b>Inicial do nome</b>
Participante 1: L1W	Líder	1	W
Participante 2: L1V	Líder	1	V
Participante 3: N1A	Não-líder	1	A
Participante 4: N1L	Não-líder	1	L
Participante 5: L2L	Líder	2	L
Participante 6: L2R	Líder	2	R
Participante 7: N2L	Não-líder	2	L
Participante 8: N2M	Não-líder	2	M
Participante 9: L3R	Líder	3	R
Participante 10: L3J	Líder	3	J
Participante 11: N3Z	Não-líder	3	Z
Participante 12: N3G	Não-líder	3	G

No QUADRO 6, o código atribuído a cada participante neste trabalho é apresentado na primeira coluna. Esse código apresenta 3 caracteres. O primeiro caractere são as letras “L” e “N”, que designam *indivíduo líder* e *indivíduo não-líder* respectivamente (ver segunda coluna do quadro). O segundo caractere são os números “1”, “2” e “3” que designam *Corpus 1*, *Corpus 2* e *Corpus 3* respectivamente (ver terceira coluna do quadro), os quais são considerados neste estudo, sendo discutidos na seção 4.4.1. O terceiro e último caractere é a inicial do primeiro nome do participante (ver quarta coluna do quadro). Assim, no QUADRO 6, o código “L1W” do primeiro participante indica que esse informante é líder (“L”), é proveniente do Corpus 1 (“1”) e tem a letra “W” como inicial de seu primeiro nome. Passemos à descrição do material desta pesquisa, o que é feito a seguir.

#### **4.4 Material**

Propõe-se aqui montar, com as 30 palavras consideradas neste trabalho, especialmente no estudo transversal, a entrevista de fala espontânea e as tarefas de leitura e de nomeação por figura, a serem aplicadas nos 12 informantes selecionados nesta pesquisa. Desse modo, o método de coleta de dados aqui utilizado é o de entrevista e o de tarefas (de leitura e de nomeação por figura). Na próxima seção, 4.4.1, são discutidos os corpora (fenômenos) considerados neste estudo. Na seção 4.4.2, a entrevista e as tarefas aqui utilizadas são brevemente descritas e, na seção 4.4.3 (de 4.4.3.1 a 4.4.3.4), são apresentadas e analisadas as 30 palavras consideradas nesta pesquisa.

##### **4.4.1 Discussão sobre os 3 corpora – 3 fenômenos do português brasileiro considerados**

Propusemos, neste trabalho, como um primeiro procedimento metodológico, a análise dos três fenômenos (Fenômenos 1 (r #), 2 (v-v) e 3 (cc))/ corpora (Corpus 1, 2 e 3) em estudo, apresentados no QUADRO 1 (Cf. capítulo 1).

Esses 3 corpora foram escolhidos devido, em especial, a três motivos. O primeiro motivo é que os fenômenos focalizados nesses corpora envolvem líquidas e estudar o comportamento das líquidas no português brasileiro é um tema de grande interesse desta pesquisadora desde o seu mestrado (FONTES MARTINS, 2001). Ainda, como se mencionou no capítulo 1, tais corpora também lidam com fenômenos que aqui

denominamos de “cancelamento”, apesar de reconhecermos que estes são casos de lenição<sup>18</sup>.

O segundo motivo é que todos os três corpora se baseiam em fenômenos do português contemporâneo de Belo Horizonte. Esta variedade do português é focalizada na presente pesquisa por facilitar a realização de nosso estudo, já que este é feito em tal cidade. Especialmente, no nosso caso em que, além da facilidade de obtenção dos dados da atual pesquisadora, Fontes Martins (2001), adotam-se dados de outras duas pesquisadoras, Cristóvão Silva (2000a, b, 2003a) e Huback (2003), as quais, juntamente com Fontes Martins, integram uma mesma equipe de pesquisa (Cf. “Diretório dos Grupos de Pesquisa no Brasil – CNPq” no *site*: <<http://lattes.cnpq.br/censo2002/>>. Acesso em: 20 ago. 2004). Deve-se dizer também que todas essas três pesquisadoras pertencem à mesma instituição de ensino superior, a Universidade Federal de Minas Gerais (UFMG). Vale mencionar, ainda, que os três fenômenos em análise parecem se encontrar em um estágio de mudança similar, o que será investigado na análise feita para tais fenômenos, no capítulo 5.

O terceiro e último motivo pelo qual adotamos os corpora 1, 2 e 3 é que os estudos de Huback (2003), Fontes Martins (2001) e Cristóvão Silva (2000a, b, 2003a) consideram a teoria da Difusão Lexical (CHEN; WANG, 1975) no tratamento dos seus dados. Como se destacou no capítulo 3, a Difusão Lexical é importante a este trabalho, sobretudo, por postular que a mudança sonora se difunde gradualmente no léxico, o que está em consonância com a Fonologia de Uso (BYBEE, 2001).

---

<sup>18</sup> Uma importante implicação de se reconhecer os fenômenos de cancelamento aqui estudados como casos de lenição é a idéia da gradualidade fonética da mudança sonora (PIERREHUMBERT, 2002). Por questões metodológicas, optamos por adotar duas possibilidades categóricas para os fenômenos em análise: cancelamento ou não-cancelamento do som (Cf. capítulos 5 e 6). Contudo, reconhecemos que existe gradualidade fonética em fenômenos de mudança e variação sonora. Como visto no capítulo 3, adotamos a noção de que a mudança sonora é lexicalmente e foneticamente gradual (BYBEE, 2001; CHENG; WANG, 1975).

#### 4.4.2 Entrevista e tarefas

As 30 palavras aqui consideradas – as quais serão apresentadas na seção 4.4.3.4 – serviram de base para a elaboração da entrevista e das tarefas a serem realizadas nesta pesquisa, mais especificamente no estudo transversal. A entrevista e as tarefas são descritas a seguir<sup>19</sup>:

- b) entrevista de fala espontânea → é apontada pela literatura sociolinguística (LABOV, 1972) como a maneira mais apropriada de se obter uma fala próxima da realidade. Para a realização dessa entrevista, foi elaborado, para servir como guia, um roteiro (APÊNDICE B) com as 30 palavras selecionadas nesta pesquisa;
  
- b) tarefa de nomeação por figura → nesta tarefa, são apresentadas figuras com desenhos de elementos que levem o informante a dizer as 30 palavras que interessam à pesquisa, por meio de perguntas que o entrevistador faz a ele. Por exemplo, para a palavra *óculos*, pode-se apresentar ao informante uma figura com este objeto e perguntar-lhe: “Como se chama esse objeto que está no rosto dessa pessoa e que é usado para enxergar melhor?”. Espera-se que o entrevistado responda: “óculos”;

---

<sup>19</sup> Na seção 4.5, a entrevista e as tarefas serão discutidas em maiores detalhes.

- c) tarefa de leitura → nesta tarefa, deverá ser feita a leitura de um texto com os 30 dados que interessam ao presente trabalho. Obviamente, tal texto contém palavras distratoras, que desviam a atenção do entrevistado da análise em foco neste estudo (Cf. APÊNDICE C).

Vale dizer que, das 30 palavras aqui consideradas, 15 são provenientes dos corpora 1, 2 e 3, considerados por esta pesquisa. Os 15 dados desses corpora foram obtidos de um único estilo de fala: todos são provenientes da entrevista de fala espontânea. Isso se deve ao fato de que esse tipo de entrevista foi o único presente em todos os 3 corpora. Assim, Cristóvão Silva (2000a, b, 2003a), além da entrevista de fala espontânea, utilizou-se das tarefas de nomeação por figura e de leitura em sua coleta de dados. Fontes Martins (2001) adotou, além da fala espontânea, a tarefa de nomeação por figura. Já Huback (2003), apesar de ter realizado tarefa de leitura, só considerou os seus dados da entrevista de fala espontânea. Como se pode perceber por essas explicações sobre a entrevista e as tarefas adotadas nos corpora 1, 2 e 3, a fala espontânea é realmente o único estilo de fala presente, ao mesmo tempo, em todos esses corpora que são considerados neste estudo.

#### **4.4.3 Apresentação e análise das 30 palavras selecionadas neste estudo**

A partir deste momento, apresentaremos e analisaremos as 30 palavras deste estudo (da seção 4.4.3.2 à seção 4.4.3.4), as quais serviram de base para a entrevista de fala espontânea e para as tarefas de leitura e de nomeação por figura do estudo transversal<sup>20</sup>.

---

<sup>20</sup> Dentre essas 30 palavras, 16 são avaliadas também no estudo longitudinal aqui realizado, o que será esclarecido no capítulo 6.

Contudo, antes de tratar propriamente dessas 30 palavras, esclarecemos, na próxima seção, algumas mudanças que ocorreram na condução deste estudo, em função de um teste-piloto aqui realizado.

#### 4.4.3.1 O teste-piloto

A fim de avaliar os dados que selecionamos inicialmente para esta pesquisa, fizemos um teste-piloto. Este foi aplicado em setembro de 2004, em Belo Horizonte, e contou com duas informantes do sexo feminino:

- a) Informante 1 → 24 anos; nascida e criada em Belo Horizonte; com nível de escolaridade de 2<sup>o</sup> grau; profissão de manicure;
- b) Informante 2 → 33 anos; nascida em Capelinha (MG), veio aos 7 anos para Belo Horizonte, onde reside até hoje; nível de escolaridade superior; estudante (mestranda em Estudos Literários/ Letras).

Nesta pesquisa, foi realizada uma primeira seleção das palavras que iriam compor a entrevista de fala espontânea e as tarefas de nomeação por figura e de leitura. Em tal seleção, escolhemos 50 palavras: as 5 palavras mais freqüentes e as 5 palavras menos freqüentes<sup>21</sup> de cada um dos corpora 1, 2 e 3, o que totaliza 30 palavras; e também as 5 palavras mais freqüentes e as 15 palavras menos freqüentes de corpora eletrônicos, LAEL-escrita e fala e CETEM/ Público<sup>22</sup>, o que totaliza mais outras 20 palavras.

---

<sup>21</sup> Deve-se ressaltar que a freqüência de ocorrência (Cf. BYBEE, 2001) é a que é considerada neste trabalho, em todos os momentos em que se menciona o vocábulo “freqüência” sem especificação.

<sup>22</sup> Os corpora eletrônicos do CETEM/ Público e do LAEL-escrita e fala serão discutidos, respectivamente, nas seções 4.4.3.2.2.1 e 4.4.3.2.2.2.

Contudo, tivemos de mudar nossos critérios de escolha de palavras, em razão de o teste-piloto aqui realizado ter demonstrado uma inaplicabilidade das palavras menos freqüentes na entrevista de fala espontânea e na tarefa de nomeação por figura, o que é explicado ainda nesta seção.

Os dados que foram avaliados no teste-piloto – 50 palavras que basearam a entrevista de fala espontânea e as tarefas de nomeação por figura e de leitura – são apresentados no QUADRO 7:

**QUADRO 7: As 50 palavras utilizadas no teste-piloto**

Corpora	Grau de frequência	Palavra	Fenômeno
<b>CORPUS 1: Huback (2003)</b>	Mais freqüente	1. Computador	<b>1. Cancelamento de “r” final em nominais</b>
		2. Lugar	
		3. Melhor	
		4. Mulher	
		5. Qualquer	
	Menos freqüente	6. Ar	
		7. Câncer	
		8. Doutor	
		9. Familiar	
		10. Lar	
<b>CORPUS 2: Fontes Martins (2001)</b>	Mais freqüente	11 Ela	<b>2. Cancelamento de líquida intervocálica</b>
		12. Ele	
		13. Eles	
		14. Óculos	
		15. Para	
	Menos freqüente	16. Cérebro	
		17. Diferenciação	
		18. Etária	
		19. Miséria	
		20. Silicone	
<b>CORPUS 3: CRISTÓFARO SILVA</b>	Mais freqüente	21. Outro	<b>3. Cancelamento de encontro consonantal tautossilábico</b>
		22. Primeiro	
		23. Problema	
		24. Quatro	
		25. Sempre	
	Menos freqüente	26. Bronzeada	
		27. Complicada	
		28. Depressão	
		29. Tranqüila	
		30. Traseira	
<b>LAEL-escrita e fala</b>	Mais freqüente	31. Maior	<b>1. Cancelamento de “r” final em nominais</b>
		32. Agora	<b>2. Cancelamento de líquida intervocálica</b>
		33. Era	<b>3. Cancelamento de encontro consonantal tautossilábico</b>
		34. Grande	
		35. Outros	
	Menos freqüente	36. Alternador	<b>1. Cancelamento de “r” final em nominais</b>
		37. Falsificador	
		38. Impostor	
		39. Incentivador	
		40. Instalador	
		41. Capivara	<b>2. Cancelamento de líquida intervocálica</b>
		42. Cítara	
		43. Filmadora	
		44. Marmelo	
		45. Versículo	
		46. Apodrecer	<b>3. Cancelamento de encontro consonantal tautossilábico</b>
		47. Braçal	
		48. Platônico	
		49. Sobreaviso	
		50. Supérflua	

Fonte: <http://www.linguateca.pt/CETEMPUBLICO/> e <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-26 de fevereiro de 2004) e Huback (2003), Fontes Martins (2001) e Cristófaros Silva (2003)

O QUADRO 7 apresenta, na terceira coluna, as 50 palavras que foram consideradas no teste-piloto realizado neste trabalho. A primeira coluna mostra o corpus de onde cada palavra foi selecionada, e a segunda coluna demonstra o grau de frequência (mais ou menos freqüente) de cada uma das 50 palavras. A última coluna discrimina o fenômeno – Fenômenos 1 (r #), 2 (v-v) e 3 (cc) – a que cada palavra está relacionada.

Como se pode perceber pela análise do QUADRO 7, há uma certa assimetria nos dados do LAEL-escrita e fala (Cf. palavras do número 36 ao 50 do QUADRO 7), em relação à frequência. Enquanto que, para cada um dos corpora 1, 2 e 3, há exatamente 5 palavras mais frequentes e 5 palavras menos frequentes, para o LAEL-escrita e fala, há 15 palavras menos frequentes, mas somente 5 palavras mais frequentes, quando se esperaria que houvesse 15 palavras mais frequentes também. Vale dizer que essa assimetria nos dados desse corpus eletrônico se deve ao fato de que tivemos de excluir dele 10 palavras, as quais coincidiam com palavras mais frequentes dos corpora 1, 2 e 3 (Cf. palavras de número 2, 3, 4, 5, 11, 12, 13, 21, 22, 24 do QUADRO 7).

É importante mencionar que as 50 palavras do QUADRO 7 foram selecionadas segundo alguns critérios. Contudo, não esclareceremos todos eles, visto que nem todas as 50 palavras serão utilizadas nesta pesquisa<sup>23</sup>. Somente serão explicitados, ainda neste capítulo, os critérios de seleção das palavras que utilizamos na nossa análise definitiva: 30 palavras mais frequentes dos corpora 1, 2 e 3 e do Corpus LAEL-escrita (Cf. QUADRO 7).

Pelo teste-piloto, pudemos constatar que as palavras menos frequentes – tanto dos corpora 1, 2 e 3 quanto dos corpora eletrônicos, LAEL-escrita e fala e CETEM/ Público – não ocorriam na entrevista de fala espontânea e na tarefa de nomeação por figura, na fala dos informantes. Tais palavras só ocorreram na tarefa de leitura. Mesmo com todo o esforço da pesquisadora em fazê-las acontecer na entrevista de fala espontânea e na tarefa de nomeação por leitura – evidentemente, sem que o informante percebesse tal esforço –, a ocorrência de palavras menos frequentes era mínima. Isso já era esperado que acontecesse, pois palavras menos frequentes apresentam menos exemplares, sendo

---

<sup>23</sup> No APÊNDICE E, é possível ter acesso à explicação dos critérios que nortearam a seleção de todas as 50 palavras que compuseram o teste-piloto.

menos acessadas no uso e tendo, portanto, representação mais fraca na memória (Cf. BYBEE, 2001; PIERREHUMBERT, 2001b).

Deve-se dizer que uma primeira dificuldade em lidar com essas palavras foi compor, com elas, perguntas para a entrevista de fala espontânea e obter figuras para a tarefa de nomeação por figura. Não foi possível compor perguntas e obter figuras para todas as palavras menos freqüentes. Dentre as palavras menos freqüentes que não ocorreram nem na entrevista, nem na tarefa de nomeação por figura, podemos citar: *alternador*, *supérflua*, *incentivador*, *diferenciação*, *complicada*, *familiar*.

Um dos aspectos que favoreceu a não-ocorrência desses exemplos de palavras menos freqüentes foi o fato de que todas essas palavras têm sentido abstrato. Isso se torna mais problemático ainda no caso da tarefa de nomeação por figura. Neste tipo de tarefa, é recomendável trabalhar-se com palavras de conteúdo, de sentido concreto.

Assim, a partir desse teste-piloto, resolvemos não considerar palavras menos freqüentes. Além do fato de que tais palavras não ocorreram no teste-piloto na fala espontânea e na tarefa de nomeação por figura, como explicitado, temos um outro motivo para não adotar palavras menos freqüentes nesta pesquisa. Palavras menos freqüentes – pelo fato de serem marcadas e chamarem mais a atenção do falante – tendem a não variar em um estilo formal (BYBEE, 2001). Assim, nesse caso, o mais provável é que o falante utiliza a variante padrão a todo o momento.

Por outro lado, se isto acontece, ou seja, se o fenômeno fonológico não ocorre em palavras de baixa freqüência – as quais foram elicitadas com técnicas que chamam a

atenção para a palavra, o que caracteriza um estilo formal – não é possível saber se o fato de o fenômeno fonológico não ter ocorrido se deve:

- a) ao estilo de fala formal;
- b) ou à baixa frequência da palavra.

Sendo assim, com essa sobreposição de razões para as palavras menos frequentes não ocorrerem, consideramos mais adequado, no caso deste trabalho, excluir tais palavras. Como se pode notar, foi importante a aplicação do teste-piloto para uma reorientação da presente pesquisa. Após tê-lo feito, reestruturamos este estudo e, com isso, realizamos uma nova seleção de dados, o que será discutido a seguir.

#### **4.4.3.2 Seleção inicial de 15 das 30 palavras em análise**

Na escolha das 30 palavras que compõem esta pesquisa, primeiramente, selecionamos as 5 palavras mais frequentes de cada um dos 3 corpora (corpora 1, 2 e 3), o que totalizou em 15 palavras. Estas são apresentadas em ordem decrescente de frequência quanto aos corpora 1, 2 e 3, na TAB. 1:

**TABELA 1: 15 palavras mais freqüentes: primeira seleção**

Corpora/ Fenômenos	Palavras	Freqüência	n e % de cancelamento
<b>Corpus 1 / Fenômeno 1 (r #):</b>  <b>Huback (2003)</b>	1. Melhor	93	24 (26%)
	2. Lugar	89	35 (39%)
	3. Mulher	74	21 (28%)
	4. Qualquer	40	28 (70%)
	5. Computador	35	16 (46%)
<b>Corpus 2 / Fenômeno 2 (v-v):</b>  <b>Fontes Martins (2001)</b>	6. Para	859	172 (20%)
	7. Ele	214	26 (12%)
	8. Ela	169	34 (20%)
	9. Eles	144	42 (29%)
	10. Óculos	40	32 (80%)
<b>Corpus 3 / Fenômeno 3 (cc):</b>  <b>Cristófarro Silva (2003)</b>	11. Problema	77	26 (34%)
	12. Outro	58	44 (76%)
	13. Quatro	57	35 (61%)
	14. Sempre	51	28 (55%)
	15. Primeiro	48	15 (31%)

Fonte: Huback (2003), Fontes Martins (2001), Cristófarro Silva (2003)

A TAB. 1 apresenta, na segunda coluna, as 15 palavras que, inicialmente, foram selecionadas para compor este trabalho. Essas 15 palavras estão distribuídas nos três corpora indicados na primeira coluna de tal tabela. A terceira coluna mostra a freqüência de ocorrência de cada uma dessas palavras nos corpora 1, 2 e 3 de onde elas foram selecionadas. A quarta coluna apresenta os números absolutos e percentuais de cancelamento das palavras no seu respectivo corpus.

As palavras mais freqüentes do Corpus 1, o qual focalizou o cancelamento de “r” final em nominais (Fenômeno 1 (r #)) foram selecionadas de Huback (2003). As palavras mais freqüentes do Corpus 2, que enfocou o cancelamento de líquidas intervocálicas (Fenômeno 2 (v-v)), foram selecionadas de Fontes Martins (2001). Já as palavras mais freqüentes do Corpus 3, que se centrou no cancelamento de líquidas em encontro consonantal tautossilábico (Fenômeno 3 (cc)), foram selecionadas de Cristófarro Silva (2003). Os critérios para a seleção dessas palavras são explicitados na seção 4.4.3.2.1 a seguir.

#### 4.4.3.2.1 Critérios de seleção das palavras da TAB. 1

É importante fazer aqui alguns esclarecimentos a respeito da seleção das 5 palavras mais freqüentes de cada um dos corpora 1, 2 e 3 (Cf. TAB. 1). De modo geral, no caso da escolha das palavras mais freqüentes dos corpora 1, 2 e 3, decidimos considerar, como um critério para indicar maior freqüência, palavras que tivessem, ao mesmo tempo, maior ocorrência e maior aplicação do fenômeno analisado.

Em relação ao ambiente estrutural mais favorável à realização do cancelamento de “r” final em nominais (Fenômeno 1 (r #) / Corpus 1), Huback (2003, p. 123) afirma que: “A princípio, parece que os sufixos *-dor* e *-tor*, independentemente de serem agentivos, estão liderando o processo de mudança (...)”.

Já no caso do cancelamento de líquidas intervocálicas (Fenômeno 2 (v-v) / Corpus 2), a posição postônica é o contexto fonético favorecedor. Como afirma Fontes Martins (2001, p. 74): “A posição postônica final pode ser apontada como um fator favorecedor à aplicação do fenômeno em estudo.” Fontes Martins (2001, p. 106) complementa: “[...] podemos dizer que o contexto postônico final é aquele em que se observa mais recorrentemente o cancelamento de líquidas intervocálicas”.

Também quanto ao cancelamento de líquidas em encontro consonantal tautossilábico (Fenômeno 3 (cc) / Corpus 3), a posição postônica favorece mais a aplicação desse fenômeno (Cf. CRISTÓFARO SILVA, 2003a). Contudo, a posição pretônica se mostra

relevante nesse caso. Isso aponta para o fato de que o ambiente átono, de modo geral, favorece o cancelamento de líquidas em encontro consonantal tautossilábico (Cf. CRISTÓFARO SILVA, 2003b).

É bom ressaltar que os três corpora (1, 2 e 3) considerados neste trabalho abordam fenômenos diferentes com metodologias distintas. Por esse motivo, apesar de o critério aqui adotado na seleção das palavras mais freqüentes ter sido o mesmo para os 3 corpora, os números encontrados para as palavras desses corpora, por diversas vezes, são muito distantes.

Um exemplo disso pode ser visto na TAB. 1 da seção anterior. De modo geral, os corpora 1 e 3 se aproximam por apresentarem valores de ocorrência parecidos para as palavras. Já o Corpus 2 se distancia desses outros dois corpora, porque os valores de ocorrência que ele apresenta para as palavras são bem maiores que os dos corpora 1 e 3. Desse modo, é preciso relativizar os valores (valor absoluto e valor percentual) de cada palavra ao seu respectivo corpus (1, 2 ou 3).

Passemos, agora, a especificar com maiores detalhes a seleção das 5 palavras mais freqüentes de cada um dos corpora 1, 2 e 3. Na escolha das 5 palavras mais freqüentes do Corpus 1, consideramos palavras que tiveram de 31 a 100 ocorrências. Dentro desse universo, selecionamos as palavras *melhor*, *lugar*, *mulher*, *qualquer* e *computador*, visto que elas apresentaram, simultaneamente, acima de 35 ocorrências e acima de 26% de aplicação do fenômeno de cancelamento do “r” final (Cf. TAB. 1 da seção 4.4.3.2). Assim, no caso do Corpus 1, o critério para a seleção de apenas 5 palavras mais freqüentes foi baseado nestes números: mais de 35 ocorrências e, ao mesmo tempo, mais de 26% de cancelamento.

Tais números não puderam ser adotados para os outros dois corpora, já que, como se apontou, por várias vezes, os valores encontrados em cada um dos corpora (1, 2 e 3) se mostraram muito distantes.

No Corpus 2, a escolha das 5 palavras mais freqüentes foi baseada em Fontes Martins (2001, p. 103) que, ao explicitar o critério de freqüência de que se utiliza, cita quais foram as 5 palavras mais freqüentes em seu corpus:

Selecionamos essas palavras [mais expressivas estatisticamente] segundo o critério de freqüência – no sentido de elas terem ocorrido mais em nosso corpus, acima de 2% de ocorrência – e também pelo maior índice de cancelamento de líquidas intervocálicas em tais palavras. Estes itens são: *ela, ele, eles, óculos e para*. (FONTES MARTINS, 2001, p. 103).

Como se pode perceber por meio dessa citação, o critério que Fontes Martins adota para encontrar as palavras mais freqüentes de seu corpus é o mesmo que adotamos aqui (Cf. primeiro parágrafo desta seção).

Quanto ao Corpus 3, as cinco palavras mais freqüentes escolhidas foram: *problema, outro, quatro, sempre e primeiro* (Cf. TAB. 1 da seção 4.4.3.2). Neste corpus, consideramos como mais freqüentes palavras que apresentaram, concomitantemente, mais de 48 ocorrências e acima de 30% de aplicação do fenômeno de cancelamento de líquidas em encontro consonantal tautossilábico.

Todas essas opções metodológicas apresentadas até o momento visaram fornecer a esta pesquisa uma seleção apropriada das palavras mais freqüentes dos corpora 1, 2 e 3, para a elaboração da entrevista e das tarefas a serem aplicados nas coletas de dados. Considerou-se importante neste trabalho fazer também uma pesquisa da freqüência

dessas palavras selecionadas em corpora eletrônicos do português brasileiro. Isso é feito na próxima seção com o intuito de se avaliar o comportamento das palavras selecionadas em corpora que são mais representativos da língua, por apresentarem um número bastante significativo de palavras.

#### **4.4.3.2.2 Avaliação, em corpora eletrônicos, dos dados obtidos dos corpora 1, 2 e 3**

Um outro procedimento metodológico adotado neste estudo foi analisar a frequência das 15 palavras da TAB. 1 (Cf. seção 4.4.3.2) nos corpora eletrônicos do CETEM/ Público (Corpus de Extratos de Textos Eletrônicos MCT/ Público) – mais especificamente, o Corpus NILC/ Universidade de São Carlos – e do LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala<sup>24</sup>. Estes corpora foram escolhidos, por terem sido, na época em que este estudo foi realizado, os melhores corpora eletrônicos disponíveis em termos de acessibilidade e quantidade de dados. Tais corpora eletrônicos são discutidos nas seções 4.4.3.2.2.1 e 4.4.3.2.2.2, e a avaliação dos 15 dados (TAB. 1 da seção 4.4.3.2), nesses corpora, é feita em 4.4.3.2.2.3.

---

<sup>24</sup> Os endereços eletrônicos do CETEM/Público e do LAEL (escrita e fala) são <<http://www.linguateca.pt/CETEMPublico/>> e <<http://LAEL.pucsp.br/corpora/index.htm>>, respectivamente. Maiores informações sobre esses dois corpora podem ser obtidas nesses endereços eletrônicos.

#### 4.4.3.2.2.1 Corpus CETEM/ Público - Corpus NILC/ Universidade de São Carlos

Dentro do corpus do CETEM/ Público, há o Corpus do NILC/ São Carlos que apresenta textos em português brasileiro, nos registros jornalístico, didático, epistolar e em redações de alunos. A TAB. 2 apresenta os dados quantitativos do Corpus NILC/ São Carlos:

**TABELA 2: Dados quantitativos do Corpus NILC/ São Carlos**

Número de frases	Número de parágrafos	Número de palavras (formas)	Número de palavras (tipos)	Número de elementos
1.964.150	827.363	32.091.996	433.030	41.372.943

Fonte: <<http://www.linguateca.pt/CETEMPublico/>> (Acesso em: 22 fev. 2004)

O corpus NILC/ São Carlos apresenta um comando que permite verificar a frequência da palavra que se deseja, bem como lista cada uma das ocorrências da palavra. Como um exemplo, apresentamos a seguir o resultado da pesquisa realizada para a palavra “freqüentador”:

**Resultado da procura pela palavra “freqüentador” no CETEM/ Público,  
Corpus NILC/ São Carlos**

**CONSULTA 1**

“Procura: **“freqüentador”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1  
2 ocorrências.

**Concordância**

Procura: **“freqüentador”**.  
2 ocorrências.

---

par 9911: De acordo com um **freqüentador** do bar, dois homens, que se identificaram como policiais, começaram a perseguir um rapaz pela Ataulfo de Paiva (na calçada do Jobi) .

---

par Informática-96a-31-40: - Não sou um assíduo **freqüentador** do encontro da gALLera<sup>25</sup>.”

É preciso dizer que o Corpus NILC/ São Carlos (e todo o restante do CETEM/ Público) oferece resultados diferentes para variações de letras maiúsculas e minúsculas, ou para variações de acento, como por exemplo, o próprio trema presente na palavra “freqüentador” apresentada acima. Isso pode ser um refinamento de pesquisa interessante para alguns trabalhos, mas veremos que não o é para este estudo.

Desse modo, esgotamos a seguir as possíveis variações de consultas que poderiam ser feitas para a palavra “freqüentador” (em relação a letras maiúsculas e minúsculas no início da palavra e em relação ao trema no “ü”), a fim de demonstrar diferentes resultados de pesquisa no Corpus NILC/ São Carlos. Note-se, contudo que, apesar de o *site* do Corpus NILC/ São Carlos listar cada uma das ocorrências da palavra pesquisada, desta vez, diferentemente da Consulta 1, não apresentamos as ocorrências uma a uma, no caso, para as Consultas 2, 3 e 4 que tiveram 52, 6 e 2 ocorrências respectivamente.

---

<sup>25</sup> Consulta da freqüência da palavra “freqüentador”: disponível no *site* do CETEM/ Público – Corpus NILC/ São Carlos (Cf. <<http://www.linguateca.pt/CETEMPUBLICO/>>. Acesso em: 25 fev. 2004).

**Resultados da procura pela palavra “freqüentador” no CETEM/ Público,  
Corpus NILC/ São Carlos**

**CONSULTA 2**

“Procura: **“frequentador”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1

52 ocorrências.

**CONSULTA 3**

Procura: **“Frequentador”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1

6 ocorrências.

**CONSULTA 4**

Procura: **“Frequëntador”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1

2 ocorrências.

**CONSULTA 5**

Procura: **“FREQUENTADOR”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1

Nenhum caso foi encontrado!

**CONSULTA 6**

Procura: **“FREQUËNTADOR”**.  
Pedido de uma concordância em contexto  
Corpus: NILC/São Carlos v. 7.1

Nenhum caso foi encontrado!

Há vantagens e desvantagens nesse detalhamento de pesquisa oferecido pelo Corpus NILC/ São Carlos. Como vantagem, podemos citar o caso de quem necessita dessa diferenciação entre letras maiúsculas e minúsculas ou o caso de quem, por exemplo, trabalhando com texto escrito, deseja observar variações (ou desvios ortográficos) relacionadas a acento gráfico.

Uma desvantagem desse refinamento de pesquisa do Corpus NILC/ São Carlos poderia ser citada, como no caso de algum estudioso que não tivesse interesse nas variações da forma escrita de uma palavra. Assim, ele estaria interessado em observar algum aspecto da palavra independente da sua grafia, por exemplo, a frequência dessa palavra na língua (sem discriminar entre escrita e fala). Para esse estudioso, a pesquisa no Corpus NILC/ São Carlos seria mais complexa, visto que ele teria de esgotar todas as possibilidades de variações na forma escrita dessa palavra para chegar ao resultado esperado.

O presente trabalho se encaixa mais neste segundo caso que é de uma desvantagem do Corpus NILC/ São Carlos, já que não estamos observando aqui fenômenos da modalidade escrita da língua. Os 3 fenômenos abordados nesta pesquisa (Cf. corpora 1, 2 e 3) ocorrem na fala<sup>26</sup>.

Como um modo de neutralizar a desvantagem dessas variações de resultados de consulta para uma mesma palavra no Corpus NILC/ São Carlos, propusemos aqui fazer a consulta das 15 palavras da TAB. 1 (seção 4.4.3.2) considerando a sua escrita padrão, dicionarizada (Cf. *Novo Dicionário Aurélio da Língua Portuguesa*, em formato eletrônico, disponível em: <<http://www2.uol.com.br/aurelio/index.html>>. Acesso em: mar. 2004). Assim, no caso da palavra “freqüentador”, somente a Consulta 1 seria feita.

Adotando esse critério, tivemos de fazer apenas uma consulta para cada palavra no Corpus NILC/ São Carlos, o que simplificou a pesquisa. Ainda, vale dizer que todas as

---

<sup>26</sup> Na realidade, marcas de oralidade podem ocorrer na escrita também. Um exemplo de estudo nesse sentido é Freitas (2001), que avalia a interferência da oralidade na escrita de crianças em fase de alfabetização, quanto ao fenômeno de cancelamento de líquida em encontro consonantal tautossilábico. Contudo, esse tipo de análise não é o foco desta pesquisa, a qual trabalha com dados de produção fala.

15 palavras da TAB. 1 (seção 4.4.3.2) foram consultadas, utilizando-se todas as letras em minúsculo e todos os acentos, como o faz o dicionário.

#### **4.4.3.2.2 Corpus LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala**

O corpus do LAEL (Linguística Aplicada e Estudos da Linguagem) também apresenta dados do português brasileiro e conta com 240 milhões de palavras (*tokens*). Contudo, no seu *site* (Cf. <<http://LAEL.pucsp.br/corpora/index.htm>>), é oferecido acesso a uma amostra do corpus de aproximadamente 1,1 milhão de palavras apenas. Neste estudo, fizemos a pesquisa de frequência de palavras no *site* do LAEL-escrita e fala e, por isso tivemos acesso a essa amostra de cerca de 1,1 milhão de palavras.

A TAB. 3 apresenta, na primeira coluna, os registros que compõem o corpus do LAEL-escrita e fala, discriminando, na terceira coluna, os tipos de textos desses registros. A segunda e a quarta colunas apresentam os totais de palavras para os registros de texto e para os tipos de texto, respectivamente.

**TABELA 3: Totais de palavras dos registros e tipos de textos do Corpus do LAEL**

<b>Registros de texto</b>	<b>Total de palavras de cada registro</b>	<b>Tipos de texto</b>	<b>Total de palavras de cada tipo de texto</b>
1. Registro acadêmico	198.652	1) Artigos e teses acadêmicas 1) Cartas comerciais 2) Cartas de pedido de emprego	198.652 19.736 14.306
2. Registros de Negócios, comerciais e técnicos	386.138	1) Editais 2) Fax Comerciais 3) Relatórios anuais de negócio 4) Manuais de informática	30.829 16.131 104.255 200.881
3. Registros falados	197.901	1) Aulas 2) Conversação	84.910 112.991
4. Registro de imprensa	199.285	1) Jornal diário, impresso	199.285
5. Registro de literatura	201.018	1) Literatura de ficção	201.018
Total geral – Corpus inteiro	1.182.994	Total geral – Corpus inteiro	1.182.994

Fonte: <<http://LAEL.pucsp.br/corpora/index.htm>> (Acesso em: 22 fev. 2004)

Note-se que os registros falados (número 3 da primeira coluna da TAB. 3) são em número muito mais reduzido do que os registros da escrita, ou seja, do que todos os outros registros: em um total de 1.182.994, há apenas 197.901 (17%) palavras de registro de fala, enquanto que as outras 985.093 (83%) palavras são de registros de escrita. É importante considerar esse dado, visto que essa diferença entre a quantidade de registros de fala e a quantidade de registro de escrita tem grande influência sobre as diferenças nos números de frequência de palavras que ocupam uma mesma posição, tanto no LAEL-escrita quanto no LAEL-fala.

Geralmente, os valores de frequência do LAEL-fala são bem inferiores aos valores de frequência do LAEL-escrita. Como exemplo, veja-se a TAB. 4 que apresenta valores de frequência do LAEL-escrita e fala para as palavras “de” e “e”, as quais ocupam a mesma posição nesse corpus: ambas as palavras são o primeiro lugar. Em termos de frequência de ocorrência, a palavra “de” é o primeiro lugar do LAEL-escrita, enquanto que a palavra “e” ocupa o primeiro lugar no LAEL-fala.

**TABELA 4: Resultados da pesquisa de frequência para as palavras “de” e “e” no LAEL-escrita e fala**

Modalidade do Corpus LAEL	Posição no corpus	Palavra	Número de ocorrências
1. Escrita	1 <sup>a</sup>	DE	1.537.460
2. Fala	1 <sup>a</sup>	E	113.061

Fonte: <<http://LAEL.pucsp.br/corpora/index.htm>> (Acesso em: 25 fev. 2004)

Como se pode observar pela TAB. 4, os valores do LAEL-fala são muito inferiores aos valores do LAEL-escrita, mesmo que se analise uma única posição no corpus. Assim, é importante relativizar essa questão da diferença que há entre os valores de frequência do LAEL-escrita e do LAEL-fala. Isso deverá ser considerado especialmente na próxima seção, 4.4.3.2.2.3, que apresenta uma análise dos 15 dados obtidos dos corpora 1, 2 e 3 (Cf. TAB. 1 da seção 4.4.3.2) nos corpora eletrônicos do CETEM/ NILC São Carlos e do LAEL-escrita e fala.

Apesar de haver essa diferenciação entre o LAEL-escrita e o LAEL-fala, ainda assim é interessante a esta pesquisa avaliar os dados nesses dois casos. Por estarmos analisando fenômenos que ocorrem na fala (Cf. corpora 1, 2 e 3 na seção 4.3), devemos considerar o corpus do LAEL-fala, mesmo que ele seja mais limitado. Ao mesmo tempo, devemos considerar também o corpus do LAEL-escrita, visto que este é bem mais representativo para a análise da frequência de palavras na língua, a qual se apresenta tanto na modalidade oral quanto na modalidade escrita.

Uma observação interessante sobre o corpus do LAEL-escrita e fala é que ele apresenta um comando que permite verificar, além do número da frequência da palavra desejada (Cf. TAB. 4), o número das frequências de todas as palavras do corpus. Isto não é possível no Corpus do CETEM/ Público – NILC/ São Carlos que, por outro lado,

apresenta todas as ocorrências da palavra pesquisada (Cf. na seção 4.4.3.2.2.1, a Consulta 1 para a palavra “freqüentador” no Corpus NILC/ São Carlos), o que o Corpus do LAEL não oferece.

#### **4.4.3.2.2.3 Análise no Corpus NILC/ Universidade de São Carlos e no Corpus LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala**

Para conferir os resultados das pesquisas de freqüência nos corpora do NILC/ São Carlos e do LAEL-escrita e fala, para 15 das 30 palavras utilizadas neste trabalho (Cf. TAB. 1 da seção 4.4.3.2), apresentamos a TAB. 5 que também mostra a freqüência destas 15 palavras nos corpora 1, 2 e 3, analisados neste estudo. É preciso dizer que tais palavras se apresentam na TAB. 5, em ordem decrescente de freqüência quanto aos corpora 1, 2 e 3.

**TABELA 5: Frequência das 15 palavras no Corpus analisado (corpora 1, 2 ou 3) e nos corpora do LAEL-escrita e fala e do NILC/ São Carlos**

Corpora / Fenômeno	Palavras	Frequência no Corpus analisado (corpora 1, 2 ou 3)	Frequência no Corpus do NILC/ São Carlos	Frequência no Corpus do LAEL-escrita	Frequência no Corpus do LAEL-fala
Corpus 1 / Fenômeno 1 Huback (2003)	1. Melhor	93	14.530	17.145	1.466
	2. Lugar	89	10.200	10.808	1.469
	3. Mulher	74	7.975	11.074	1.018
	4. Qualquer	40	13.725	13.486	1.058
	5. Computador	35	2.926	3.042	99
Corpus 2 / Fenômeno 2 Fontes Martins (2001)	6. Para	859	297.628	353.847	3.611
	7. Ele	214	45.190	71.290	23.104
	8. Ela	169	16.903	25.361	13.126
	9. Eles	144	12.767	20.136	14.145
Corpus 3 / Fenômeno 3 Cristófaró Silva (2003a)	10. Óculos	40	441	627	22
	11. Problema	77	8.920	9.032	1.631
	12. Outro	58	14.926	17.686	3.549
	13. Quatro	57	15.028	18.337	2.103
	14. Sempre	51	13.667	15.513	5.690
	15. Primeiro	48	20.011	23.822	1.641

Fonte: <http://www.linguateca.pt/CETEMPublico/> e <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-30 mar. 2004) e Huback (2003), Fontes Martins (2001) e Cristófaró Silva (2003)

Esta pesquisa de frequência nos corpora do LAEL-escrita e fala e do NILC/ São Carlos, para 15 das 30 palavras utilizadas neste trabalho, é interessante por fornecer uma visão mais ampla da frequência dessas palavras na língua. É sabido que uma palavra pode ser muito frequente em um corpus particular, mas não ter grande frequência na língua. Certamente que corpora substanciais, como o LAEL-escrita e fala e o NILC/ São Carlos, podem oferecer parâmetros de frequência mais próximos da realidade da língua.

Esse é o motivo maior pelo qual propusemos a análise feita na TAB. 5 que repete, da primeira até a terceira coluna, dados da TAB. 1. Contudo, a TAB. 5 apresenta ainda, como se disse, os valores de frequência das 15 palavras dos corpora 1, 2 e 3, nos corpora do LAEL-escrita e fala e do NILC/ São Carlos. Isso é feito da quarta à sexta coluna da TAB. 5.

Como se pode ver nesta tabela, os resultados de frequência das 15 palavras dos corpora 1, 2 e 3 em relação aos corpora eletrônicos do LAEL-escrita e fala e do NILC/ São Carlos são diferenciados em alguns casos. Há casos em que os valores se mostram semelhantes em todos os corpora (corpora 1, 2 e 3, corpus LAEL-escrita e fala e corpus NILC/ São Carlos). Isso ocorre, por exemplo, com as palavras de número 1 e 2 (Cf. TAB. 5), em que todos os corpora apresentam valores mais altos para essas palavras.

Ao mesmo tempo, há também casos em que uma única palavra apresenta um valor de frequência alto em um corpus e em outro corpus não. Este é o caso da palavra de número 5 (Cf. TAB. 5), *computador*, que apresenta um valor de frequência mais alto no Corpus 1 em análise e valores de frequência relativamente mais baixos nos corpora do LAEL-escrita e fala e do NILC/ São Carlos.

É possível, no entanto, que esses baixos valores para a palavra *computador* se devam ao fato de que tal palavra é relativamente recente na língua, e os corpora eletrônicos aqui considerados abarcam também dados mais antigos. A palavra em questão, *computador*, passou a ser mais empregada na nossa língua, a partir da década de noventa, com o início da popularização do computador no Brasil. Contudo, pode-se afirmar, pela avaliação da TAB. 5, que, de modo geral, os valores de frequência encontrados nos corpora 1, 2 e 3 são confirmados pelos corpora eletrônicos do LAEL-escrita e fala e do NILC/ São Carlos.

#### 4.4.3.3 Apresentação dos outros 15 dados (obtidos do LAEL-escrita e fala)

Adicionalmente à informação dos valores de frequência – nos corpora 1, 2 e 3 e nos corpora do LAEL-escrita e fala e do NILC/ São Carlos – das 15 palavras apresentadas na TAB. 5, também selecionamos no LAEL-escrita, as 5 palavras mais frequentes para cada um dos 3 fenômenos contemplados pelos corpora 1, 2 e 3. Assim, a nova análise contou também com um total de 15 palavras: 5 palavras com “r” final em nominais, 5 palavras com líquida intervocálica e 5 palavras com encontro consonantal tautossilábico.

Escolhemos o LAEL-escrita como base para a seleção das outras 15 palavras em análise, por este apresentar um número de dados representativo e, por isso, ser um bom indicador da frequência de palavras na língua. Como se mostrou em 4.4.3.2.2.2, o LAEL-escrita apresenta um número maior de dados do que o LAEL-fala, por isso este não foi escolhido. A seleção dessas outras 15 palavras não pôde ser feita para o Corpus do NILC/São Carlos, visto que, neste corpus, não é possível ver as frequências de todas as suas palavras, uma a uma, o que é possível no LAEL-escrita e fala. No corpus do NILC/ São Carlos, só é possível fazer consulta da frequência de palavras pré-selecionadas, indicadas pelo usuário desse corpus no *site*. As outras 15 palavras mais frequentes do corpus do LAEL-escrita são apresentadas na TAB. 6, na ordem decrescente de frequência em que ocorreram neste corpus:

**TABELA 6: As outras 15 palavras mais freqüentes do LAEL-escrita, distribuídas de acordo com os fenômenos dos corpora 1, 2 e 3**

Fenômenos	Palavras	Valores de freqüência no LAEL-escrita
<b>1. Cancelamento de “r” final em nominais (Corpus 1)</b>	1. Maior	24.534
	2. Mar	13.364
	3. Governador	9.589
	4. Jogador	8.477
	5. Dólar	8.438
<b>2. Cancelamento de líquida intervocálica (Corpus 2)</b>	6. Era	32.247
	7. Dinheiro	27.555
	8. Feira	20.334
	9. Agora	19.373
<b>3. Cancelamento de encontro consonantal tautossilábico (Corpus 3)</b>	10. Lula	16.385
	11. Brasil	66.925
	12. Presidente	42.744
	13. Três	29.800
	14. Grande	24.914
	15. Livro	11.062

Fonte: <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-26 de fevereiro de 2004) e Huback (2003), Fontes Martins (2001) e Cristófarro Silva (2003)

A TAB. 6 apresenta, na segunda coluna, as 15 palavras mais freqüentes (5 palavras para cada um dos 3 fenômenos aqui analisados) retiradas do LAEL-escrita. A seleção dessas palavras considerou os fenômenos focalizados pelos corpora 1, 2 e 3, o que pode ser visto na primeira coluna. A terceira e última coluna apresenta os valores de freqüência das 15 palavras no LAEL-escrita.

Vale dizer que a seleção dessas outras 15 palavras foi relativamente simples. O *site* do LAEL disponibiliza esse utilitário com fácil acessibilidade. Maiores esclarecimentos sobre a seleção das palavras mais freqüentes são dados a seguir, na seção 4.4.3.3.1.

#### 4.4.3.3.1 Especificidades na seleção das palavras mais freqüentes do LAEL-escrita

Ao selecionar, no LAEL-escrita, as outras 15 palavras em estudo, adotamos dois critérios que passam a ser descritos. Em primeiro lugar, excluimos de nossa seleção no LAEL-escrita, palavras que coincidissem com as primeiras 15 palavras retiradas dos corpora 1, 2 e 3 (Cf. TAB. 1 da seção 4.4.3.2). Assim, tivemos de desconsiderar nessa seleção do LAEL-escrita, as seguintes 10 palavras mais freqüentes que já foram consideradas anteriormente (Cf. TAB. 1 da seção 4.4.3.2):

- quanto ao Fenômeno 1 (r #) – (Corpus 1): *melhor, qualquer, mulher e lugar*;
- quanto ao Fenômeno 2 (v-v) – (Corpus 2): *ele, ela e eles*;
- quanto ao Fenômeno 3 (cc) – (Corpus 3): *primeiro, quatro e outro*.

O segundo critério que adotamos foi selecionar, preferencialmente, palavras de conteúdo e de sentido concreto. Isso foi feito com a intenção de se garantir que todas as palavras em estudo sejam pronunciadas na entrevista de fala espontânea e na tarefa de nomeação por figura em especial. Como se disse na seção 4.4.3.1, palavras de conteúdo e de sentido concreto são mais fáceis de serem elicitadas.

#### 4.4.3.4 Apresentação geral das palavras consideradas nesta pesquisa

Este estudo contará com 30 palavras no total, as quais são apresentadas no QUADRO 8:

**QUADRO 8: As 30 palavras consideradas nesta pesquisa**

<b>Corpora</b>	<b>Palavras</b>	<b>Fenômeno</b>
<b>CORPUS 1:</b> <b>Huback (2003)</b>	1. Melhor 2. Lugar 3. Mulher 4. Qualquer 5. Computador	<b>1. Cancelamento de “r” final em nominais</b>
<b>CORPUS 2:</b> <b>Fontes Martins (2001)</b>	6. Para 7. Ele 8. Ela 9. Eles 10. Óculos	<b>2. Cancelamento de líquida intervocálica</b>
<b>CORPUS 3:</b> <b>Cristófarro Silva (2003)</b>	11. Problema 12. Outro 13. Quatro 14. Sempre 15. Primeiro	<b>3. Cancelamento de encontro consonantal Tautossilábico</b>
	16. Maior 17. Mar 18. Governador 19. Jogador 20. Dólar	<b>1. Cancelamento de “r” final em nominais</b>
<b>LAEL -escrita</b>	21. Era 22. Dinheiro 23. Feira 24. Agora 25. Lula	<b>2. Cancelamento de líquida intervocálica</b>
	26. Brasil 27. Presidente 28. Três 29. Grande 30. Livro	<b>3. Cancelamento de encontro consonantal Tautossilábico</b>

Fonte: <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-26 de fevereiro de 2004) e Huback (2003), Fontes Martins (2001) e Cristófarro Silva (2003)

O QUADRO 8 apresenta, na segunda coluna, as 30 palavras a serem consideradas, especialmente no estudo transversal do presente trabalho. A primeira coluna mostra o corpus de onde cada palavra foi selecionada, e a última coluna discrimina o fenômeno a que cada palavra está relacionada. Essas palavras servirão de base para a entrevista de fala espontânea e as tarefas de nomeação por figura e de leitura. Maiores detalhes sobre os dados que serão utilizados na entrevista e nas tarefas são dados a seguir, na seção 4.5.

#### 4.5 Procedimentos de coleta e avaliação dos dados

Montamos, com as 30 palavras consideradas neste estudo (Cf. QUADRO 8 da seção anterior) a entrevista de fala espontânea e as tarefas de nomeação por figura e de leitura, as quais foram aplicadas nos 12 informantes recontactados, quando da coleta feita para o estudo transversal. Como foi citado na seção 4.2, neste trabalho, foram realizados dois tipos de coleta: uma coleta com todos os 12 informantes para um estudo transversal e outras 6 coletas mensais com 2 desses 12 informantes para um estudo longitudinal. Assim, as 30 palavras selecionadas, adotadas na entrevista de fala espontânea e nas tarefas de nomeação por figura e de leitura, foram aplicadas na coleta feita para o estudo transversal. Já nas 6 coletas feitas para o estudo longitudinal, a princípio, foram considerados todos os dados referentes aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) discutidos anteriormente (Cf. seção 4.4.1). Contudo, no decorrer da análise, foi feito um recorte de tais dados, de modo que foram considerados somente aqueles referentes a 16 das 30 palavras avaliadas no estudo transversal, o que será esclarecido no capítulo 6.

É preciso dizer que, já com os dados da coleta transversal, pudemos fazer uma certa reavaliação desses 12 indivíduos. Isso porque, em posse também dos dados de tais informantes nas pesquisas anteriores aqui consideradas (Corpus 1: HUBACK, 2003; Corpus 2: FONTES MARTINS, 2001; Corpus 3: CRISTÓFARO SILVA, 2003), realizamos uma análise comparando o desempenho desses 12 indivíduos no presente estudo com o desempenho deles nessas pesquisas anteriores. Tal análise que tem, assim, um caráter de estudo longitudinal – por comparar a produção lingüística dos mesmos indivíduos em tempos diferentes – é apresentada na seção 6.2 do capítulo 6. Essa reavaliação é importante por nos permitir investigar se há, além de variações

interindividuais, alguma variação intra-individual entre a coleta transversal proposta neste estudo e a coleta dessas pesquisas anteriores referentes aos corpora 1, 2 e 3.

Também o estudo longitudinal proposto neste trabalho, com 6 coletas mensais, permite avaliar tais variações no comportamento dos indivíduos em análise. Vale ressaltar que os dados relativos aos estudos transversal e longitudinal são apresentados, analisados e discutidos nos capítulos 5 e 6, respectivamente. A seguir, discutimos a coleta do estudo transversal na seção 4.5.1 e as coletas do estudo longitudinal na seção 4.5.2.

#### **4.5.1 Coleta do estudo transversal**

##### **4.5.1.1 Alguns esclarecimentos a respeito da coleta do estudo transversal**

Antes de a entrevista e as tarefas da coleta transversal serem realizadas, foram dadas ao participante as instruções sobre o que ocorreria e como este deveria se portar durante a coleta. Assim, a pesquisadora esclareceu ao informante a ordem dos acontecimentos nas coletas: em primeiro lugar, a entrevista de fala espontânea; em segundo, a tarefa de nomeação por figura; e em terceiro e último, a tarefa de leitura.

Pelo fato de o estudo transversal ser feito com base na entrevista de fala espontânea e em dois tipos de tarefas (de nomeação por figura e de leitura), é possível a esta pesquisa avaliar o fator estilo de fala. Como se pode notar, há uma espécie de gradação de estilo de fala que vai de um estilo informal com a entrevista de fala espontânea, passando por

um estilo intermediário (contudo, mais para formal) com a tarefa de nomeação por figura, chegando a um estilo formal com a tarefa de leitura. Sangster (2002) sugere que o comportamento do indivíduo, especialmente quanto a variações intra-individuais, pode ser influenciado pelo estilo de fala.

Quanto ao objetivo da pesquisa, o participante foi informado apenas que a coleta tinha um caráter acadêmico. Não foi especificado o objeto de estudo, como é indicado pela literatura (LABOV, 1972), a fim de que o informante não policiasse a sua fala, utilizando uma linguagem próxima do padrão, sem variações. Ainda, ao participante, foi solicitado que se comportasse da maneira mais à vontade e natural possível, para que a coleta fosse bem sucedida.

Vale ressaltar que as gravações do estudo transversal foram realizadas na casa ou no local de trabalho dos informantes, de acordo com a disponibilidade e a preferência de cada um deles. Por esses ambientes serem familiares aos informantes, esperávamos que a fala deles apresentasse, conseqüentemente, maior naturalidade. Ainda, as gravações com cada indivíduo tiveram a duração média de 1 hora e 10 minutos, o que totaliza, aproximadamente, 13 horas e 20 minutos de gravação na coleta transversal.

Nessa coleta, além de os 12 indivíduos serem analisados em relação ao fenômeno do corpus de que fazem parte (corpora 1, 2 ou 3), eles são observados quanto aos outros dois fenômenos dos outros dois corpora, em que não foram avaliados. Assim, um informante do Corpus 1 que foi avaliado somente em relação ao cancelamento de “r” final em nominais (Fenômeno 1 (r #)), na presente pesquisa, será também avaliado quanto ao cancelamento de líquidas intervocálicas (Fenômeno 2 (v-v) / Corpus 2) e quanto ao cancelamento de líquidas em encontro consonantal tautossilábico (Fenômeno

3 (cc) / Corpus 3). Isso nos permitirá observar se o comportamento de líder ou não-líder é constante ou não em relação a fenômenos diferentes. Espera-se que a análise de fenômenos variados permita-nos delinear melhor o perfil dos indivíduos – a fim de que possamos avaliar o comportamento do indivíduo e, inclusive, a questão da liderança – e ofereça-nos elucidação quanto às teorias de representações múltiplas e à organização do componente fonológico (BYBEE, 2001; PIERREHUMBERT, 2001b; BOD; HAY; JANNEDY, 2003).

#### **4.5.1.2 Dados**

Como se disse anteriormente, a coleta feita para o estudo transversal conta com a entrevista de fala espontânea e as tarefas de nomeação por figura e de leitura. Assim, nessa coleta, como estamos trabalhando com 30 palavras (Cf. QUADRO 8 da seção 4.4.3.4) a serem obtidas em 3 formas de coleta de dados (entrevista de fala espontânea, tarefa de nomeação por figura e tarefa de leitura) com os 12 informantes aqui considerados (Cf. QUADRO 6 da seção 4.3), esperamos obter um mínimo de 1.080 dados. Os dados da coleta transversal serão apresentados, analisados e discutidos no capítulo 5.

### **4.5.1.3 Conteúdo da entrevista de fala espontânea e das tarefas de nomeação por figura e de leitura**

#### **4.5.1.3.1 Entrevista de fala espontânea**

Para a realização da entrevista de fala espontânea, foi elaborado um roteiro de perguntas, contendo as 30 palavras em análise. Isso foi feito com o intuito de se ter uma maior garantia de que estas palavras fossem pronunciadas pelo informante. Tal roteiro pode ser conferido no APÊNDICE B.

#### **4.5.1.3.2 Tarefa de nomeação por figura**

Na tarefa de nomeação por figura, foram apresentadas figuras (Cf. LADEFOGED, 2003, p. 24) com desenhos/ fotos de elementos que levassem o informante a dizer as palavras que interessavam à pesquisa, por meio de perguntas que a entrevistadora fez a ele. Assim, por exemplo, ao apresentar uma foto do atual governador do Estado de Minas Gerais, Aécio Neves, a pesquisadora perguntava ao participante quem era aquele da foto, o que ele fazia, qual era o seu cargo na política, e o participante dizia a palavra *governador* que é uma das 30 palavras focalizadas no estudo transversal. No caso dessa tarefa de nomeação por figura, geralmente cada palavra foi dita de forma isolada pelo informante, e não em um contexto de frase.

#### 4.5.1.3.3 Tarefa de leitura

A tarefa de leitura foi baseada em um texto. A nossa proposta de utilizar, nessa tarefa, um texto maior com todas as palavras em estudo (e não uma lista com essas palavras sozinhas ou em frases) teve como base a preocupação de distrair os participantes quanto aos objetivos desta pesquisa. Ainda, a leitura de um texto como o elaborado neste estudo – narrativo, com princípio, meio e fim (ver a seguir) – colabora para que a leitura seja mais natural, fluente e não tão mecânica como a leitura de palavras isoladas ou mesmo como a leitura de palavras em frases. Outra preocupação que tivemos foi a de fazer um texto com um tamanho representativo, com todas as palavras em análise nesta pesquisa, mas não tão grande a ponto de fatigar os participantes na leitura.

É preciso dizer que o texto elaborado para a tarefa de leitura apresentava duas ocorrências para cada uma das 30 palavras em foco (10 palavras para cada um dos Fenômenos 1 (r #), 2 (v-v) e 3 (cc), em análise), o que totalizou 60 ocorrências (20 ocorrências para cada um dos Fenômenos 1 (r #), 2 (v-v) e 3 (cc)). Optamos por adotar essas duas ocorrências para cada uma das 30 palavras, para termos uma maior garantia de que todas elas fossem lidas. Às vezes, ocorre de algum participante, na leitura, saltar alguma palavra que pode estar sendo focalizada pela pesquisa. Assim, ao colocarmos duas ocorrências para cada palavra em análise, minimizamos a possibilidade de o participante não pronunciá-la.

O texto utilizado como material para a leitura, “As contadoras de histórias de Tempo Feliz”, é apresentado, a seguir, com as 60 ocorrências das 30 palavras sublinhadas.

Deve-se ressaltar que todas as 60 ocorrências foram consideradas na análise e que todos os participantes leram este mesmo texto com essas ocorrências.

### Texto utilizado na tarefa de leitura

#### As contadoras de histórias de Tempo Feliz

Dona Glória e Dona Marta, duas irmãs gêmeas de oitenta e seis anos, são contadoras de histórias em um lugar pacato no norte de Minas chamado Tempo Feliz. Em geral essas irmãs passam a maior parcela do seu tempo treinando a dramatização das narrativas. Elas dizem que essa é a melhor tarefa que fazem na vida. Na opinião de todos, os contos são sempre perfeitos e têm um enredo muito envolvente. Antes de serem contadoras de histórias, a Dona Glória era tenista e a Dona Marta era protética e também dentista. Mas as duas já tiveram outros empregos antes ainda, sendo que o primeiro trabalho de Dona Glória foi de secretária e o primeiro trabalho de Dona Marta foi de professora. Depois de aposentarem é que elas descobriram esse dom delas de contar histórias. As duas irmãs têm o maior talento para contar “causos”, o que é comprovado pelo fato de que em toda e qualquer parada delas em algum lugar parece haver uma festa entre as pessoas que ouvem as histórias que elas contam.

O público é que escolhe o que elas vão contar. Dona Glória e Dona Marta atendem geralmente a três pedidos de histórias em cada sessão. As sessões ocorrem onde elas moram na Rua Quatro Paredes pelo menos duas tardes por semana. Todos ficam encantados com as histórias principalmente porque o grande conjunto delas são “causos” conhecidos pelas pessoas da cidade. Mas as duas irmãs às vezes contam coisas que acontecem agora também no nosso Brasil tachado de país de terceiro mundo por causa do problema político-econômico e social que está sempre perante a nossa população sofredora. Por exemplo, elas já estão tratando do Mensalão que é considerado como o problema político mais sério que já ocorreu nesse nosso grande país de gente humilde mas trabalhadora.

Só que Dona Glória e Dona Marta inventam coisas em cima desses acontecimentos atuais. Elas contam uma história em que até um jogador parrudo de futebol está envolvido no caso do Mensalão. De acordo com elas, o que acontece é que esse tal jogador parrudo atua muito bem em um dia mas muito mal em outro porque faz o placar que o político quiser em qualquer partida em troca de um Mensalão.

As irmãs não poupam nem mesmo o Lula contando que esse Presidente casado com o compromisso de fazer o bem à nossa nação supostamente e o Governador calado e indiferente do Acre estão envolvidos no caso do Mensalão. Nessa história elas contam que o Lula concede ilegalmente um alto valor em dinheiro parado na conta da presidência ao tal Governador também do partido desse Presidente querido pelo povo por ter sido um operário que agora pertence ao pequeno grupo que dirige o país. Dona Glória e Dona Marta criticam esses políticos dizendo que eles teriam de devolver essa quantia ao povo em forma de emprego. Mas quando as irmãs terminam essa história, elas se desculpam com eles pedindo ao público que não leve a sério a história delas que é inventada. O público aplaude entusiasmado. Todo mundo da cidade de Tempo Feliz e do Brasil parece estar abismado com esse caso do Mensalão e é por isso que ele talvez tem esse ibope enorme entre as pessoas que ouvem as histórias das irmãs.

Dona Glória e Dona Marta ainda escrevem e declamam poemas. Elas têm até um poema moderno que fala de um computador capaz de dominar os homens. A história desse computador capaz de mandar nos homens provoca muita risada entre as pessoas e alguns pensam que ela permite pensar que é possível acontecer o dia em que a máquina vai substituir o homem. Mas ainda assim nenhum outro poema é mais interessante do que o que fala do caso da mulher pagã que trabalha na Feira Cativa perto do Mar Tainha vendendo um livro pesado que se chama “O óculos preto que faz enxergar o futuro”. O mais engraçado é que as irmãs contam que esse livro pesado que fala de um óculos preto que consegue fazer previsões do futuro é vendido em dólar parado em conta bancária. Segundo elas, se o cliente não tiver dólar parado na conta bancária não é possível vender o produto mesmo porque é só pela internet que essa mulher pagã realiza a venda. Assim não adianta o freguês ir na Feira Cativa próxima ao Mar Tainha com dinheiro querendo comprar a obra. A feirante só vende esse famoso produto pela internet. O detalhe é que ela também vende três produtos apenas por freguês. Uma pessoa que quis levar quatro produtos teve de pedir ao seu melhor parente e amigo para comprar a obra como se ele quisesse comprar e não essa pessoa. Quem quiser saber mais detalhes sobre essa e outras histórias, procure Dona Glória e Dona Marta, contadoras de histórias da cidade de Tempo Feliz.

Quando da elaboração do texto “As contadoras de histórias de Tempo Feliz” apresentado na página anterior, preocupamo-nos em produzir um texto que fosse diferente, até mesmo engraçado, com a intenção, novamente, de distrair o participante quanto aos objetivos desta pesquisa. Assim, como se pode perceber, o texto apresenta um enredo que têm um efeito de humor. Como os dados desse texto mesclam elementos da realidade (por exemplo, a presença de políticos brasileiros) com ficção (por exemplo, a existência de um “óculos preto que consegue fazer previsões do futuro”), isso o torna um texto divertido, pitoresco. Prova disso é que, durante a sua leitura, os participantes, por diversas vezes, riam achando a história engraçada, segundo eles mesmos relataram durante a coleta.

Outro aspecto do texto que merece ser comentado é a pontuação. Esta não segue exatamente a norma culta. Em algumas partes do texto, segundo essa norma, deveria haver uma vírgula ou mesmo um ponto final, mas não os colocamos, a fim de diminuir as pausas e assim tornar a leitura mais fluente.

Na versão desse texto aplicada na coleta de dados, obviamente, as palavras em análise não foram sublinhadas, ou seja, todas as 30 palavras/ 60 ocorrências ficaram com o mesmo aspecto do restante das palavras do texto, sem qualquer grifo (ver APÊNDICE C). Ainda, diferentemente do texto apresentado acima (que ocupa pouco menos de uma página e tem formatação: fonte Times New Roman, número 10, espaço entre linhas simples), no texto apresentado aos informantes, a formatação empregada visou deixar o texto em um tamanho maior (com o texto ocupando duas páginas e com a seguinte formatação: fonte Arial, número 12, espaço entre linhas de 1,5). Fizemos isto com o objetivo de o texto ser melhor visualizado (ver APÊNDICE C) pelos participantes, quando da sua leitura.

Quanto a aspectos mais lingüísticos do texto, devemos ressaltar que tivemos algumas preocupações, dentre elas, a de observar o contexto fonético adjacente (especialmente, o início da palavra seguinte) a cada uma das 30 palavras/ 60 ocorrências em análise. Assim, na elaboração do texto “As contadoras de histórias de Tempo Feliz”, estabelecemos alguns critérios<sup>27</sup>, a fim de obtermos o melhor contexto lingüístico possível para cada ocorrência em foco. Tais critérios que nortearam a elaboração desse texto são apresentados e discutidos a seguir:

1. no texto, dispusemos as 30 palavras a uma certa distância uma da outra, a fim de que, na pronúncia de cada uma delas, uma não influenciasse a outra. Assim, por exemplo, distanciamos as palavras “Presidente” e “Lula” que até poderiam vir juntas (como, “Presidente Lula”), para que a pronúncia de uma não afetasse a da outra;
2. também, dispusemos as 30 palavras em análise mais no meio da frase entoacional, com o objetivo de garantir a sua pronúncia integral. Quando uma palavra ocorre no fim da frase entoacional (LADEFOGED, 2003), essa palavra tende a ter os seus sons finais comprometidos por algum processo de redução ou, até mesmo, ela pode ser apagada inteiramente;
3. quanto às 10 palavras/ 20 ocorrências com “r” final em nominais (referentes ao Fenômeno 1 (r #)), colocamo-las seguidas, sempre, de uma palavra começando com uma oclusiva desvozeada e vogal [a] em sílaba átona. Como exemplo disso, podemos citar a expressão “dólar parado” presente no texto utilizado para a tarefa de leitura. Nessa expressão, o vocábulo com “r” final em análise, “dólar”, é seguido da palavra “parado” que começa com uma consoante oclusiva desvozeada bilabial e vogal [a] em sílaba átona. Estabelecemos esse critério de adotar uma consoante

---

<sup>27</sup> Agradecemos à Prof<sup>a</sup>. Dra. Adriana Marusso (Universidade Federal de Ouro Preto – UFOP) por colaborar na elaboração do texto da tarefa de leitura, especialmente na formulação dos critérios utilizados para a composição de tal texto.

oclusiva desvozeada, a fim de evitar que o “r” final fosse pronunciado, por exemplo, como um tepe, se esse “r” fosse seguido de uma vogal (como em: “dóla[r] americano”). Ladefoged (2003, p. 7-8) propõe que se apresente a palavra analisada na sentença em um contexto fonético que favoreça a sua pronúncia integral, a realização de todos os sons que a compõem, evitando fenômenos que a modifiquem, como o de juntura (regra de Sândi), o qual pode ocorrer em, “dóla[r] americano”. A vogal [a] depois da consoante oclusiva desvozeada (por exemplo, em “p[a]rado”) também foi utilizada para garantir um contexto mais homogêneo para todas as 10 palavras/ 20 ocorrências com “r” final em nominais, focalizadas nesta pesquisa;

4. quanto às 10 palavras/ 20 ocorrências com as líquidas /l/ e /r/ intervocálicas (referentes ao Fenômeno 2 (v-v)), colocamo-las em padrão proparoxítono. Na verdade, dessas 10 palavras, precisamos colocar somente 9 que eram paroxítonas no padrão proparoxítono, já que 1 palavra, “óculos”, já era proparoxítona (Cf. QUADRO 8 da seção 4.4.3.4). Assim, optamos por colocar todas as 10 palavras em um padrão semelhante, no padrão proparoxítono, já que não seria possível fazer a palavra “óculos” tomar o padrão paroxítono, sem que ela sofresse redução no seu interior. Como exemplo desse procedimento de colocar todas as palavras no padrão proparoxítono, podemos citar a palavra paroxítona “agora” (que apresenta a seqüência: sílaba átona + sílaba tônica + sílaba átona), a qual foi seguida de uma palavra que apresentava, em seu início, a seqüência: “sílaba átona + sílaba tônica”. Assim, na sua primeira ocorrência, a palavra “agora” foi seguida da palavra “também” (sílaba átona + sílaba tônica) e, na sua segunda ocorrência, “agora” foi seguida da palavra “pertence” (sílaba átona + sílaba tônica...). Esse procedimento foi realizado para todas as demais 8 palavras paroxítonas referentes ao Fenômeno 2 (v-v). A palavra “óculos” que era a única que já apresentava o padrão proparoxítono, para preservar tal padrão, foi seguida, nas suas duas ocorrências, de uma palavra –

no caso, da palavra “preto” – com a seqüência, “sílabo tônica + sílabo átona” (conferir no texto acima, “As contadoras de histórias de Tempo Feliz”);

5. quanto às 10 palavras/ 20 ocorrências com encontro consonantal tautossilábico (referentes ao Fenômeno 3 (cc)), aquelas que apresentavam tal encontro no fim de palavra (*outro, quatro, sempre e livro*, Cf. QUADRO 8 da seção 4.4.3.4) foram seguidas de uma palavra com consoante oclusiva desvozeada em sílabo átona. Um exemplo é a palavra “quatro” que, na sua primeira ocorrência no texto acima utilizado na tarefa de leitura, é seguida da palavra “parede”. A palavra “três”, único caso de monossílabo tônico (Cf. QUADRO 8 da seção 4.4.3.4), também seguiu esse mesmo critério, por isso, na sua primeira ocorrência, foi seguida de “pedidos” e, na sua segunda ocorrência, foi seguida de “produtos”.

A seguir, na próxima seção, apresentamos e discutimos as variáveis analisadas no estudo transversal.

#### **4.5.1.3.4 Variáveis analisadas**

No estudo transversal (assim como no longitudinal), analisa-se a variável dependente *Cancelamento*, que apresenta duas categorias: (1) cancelamento, (2) não-cancelamento.

Ainda, com base nos objetivos e no suporte teórico deste trabalho, propomos, no estudo transversal, a análise de 5 variáveis independentes<sup>28</sup> que foram aqui controladas: *Fenômeno, Palavra, Estilo, Liderança e Indivíduo*. Estas são tratadas a seguir.

### 1. Fenômeno

Em estudos sobre mudança sonora, é comum analisar indivíduos com relação a apenas um fenômeno (MOTA, 2002; LEITE, 2006; TOMAZ, 2006). Desse modo, não é comum a análise de mais de um fenômeno para os mesmos indivíduos, o que aqui fazemos ao avaliarmos os Fenômenos 1 (r #), 2 (v-v) e 3 (cc). Como discutido no capítulo 2, Sangster (2002) encontrou variações inter e intra-individuais ao analisar diversos fenômenos do inglês. Assim, neste trabalho, ao considerar a variável *Fenômeno*, buscamos caracterizar variações no comportamento do indivíduo quanto a diferentes fenômenos fonológicos do português brasileiro.

### 2. Palavra

Tendo em vista que, neste estudo, consideramos a teoria da Difusão Lexical que, por sua vez, é considerada pelas teorias multirrepresentacionais em análise, avaliamos o fator palavra. Analisamos a variabilidade individual com relação a itens específicos.

---

<sup>28</sup> Inicialmente, na análise estatística, avaliamos a variável dependente *Cancelamento* com relação a 22 variáveis independentes, sendo que 17 delas foram desconsideradas completamente da presente análise, por não terem apresentado resultado relevante. Dessas 17 variáveis ignoradas, como variáveis sociais tradicionais em uma análise variacionista, tínhamos: 1. *Sexo*; 2. *Faixa etária*; 3. *Escolaridade*; e 4. *Classe social*. As outras variáveis independentes desconsideradas foram: 5. *classe gramatical*, 6. *tamanho da palavra (se monossílabo, dissílabo...)*; 7. *tonicidade*; 8. *vogal anterior*; 9. *vogal posterior*; 10. *1ª consoante do encontro consonantal tautossilábico*; 11. *tipo de líquida – “l” ou “r”, e “r” final*; 12. *estrutura silábica (CV, CCV, CVC...)*; 13. *sentido*; 14. *posição (se o som estava no início, no meio ou no fim da palavra)*; 15. *grau de empatia da pesquisadora com o informante*; 16. *local de coleta – casa ou trabalho do informante* e 17. *informante – conhecido ou não da pesquisadora*.

As 5 variáveis efetivamente consideradas neste trabalho, quando submetidas a análises estatísticas, apresentaram resultados que merecem ser aqui discutidos. Dessa forma, no próximo capítulo em que tratamos dos dados do estudo transversal, além de analisarmos as 5 variáveis separadamente, também, na análise das variáveis *Liderança e Indivíduo*, avaliamos a interação de cada uma destas com as outras variáveis estudadas: *Fenômeno, Palavra e Estilo*.

Como consideramos que a organização do léxico (e, conseqüentemente, da gramática) varia de indivíduo para indivíduo (HOPPER, 1987; BYBEE, 2001), esperamos encontrar que falantes diferentes lidam de forma diversa com as mesmas palavras. Propomo-nos a verificar essa variabilidade.

### 3. *Estilo*

A variável *Estilo*, muito analisada em trabalhos de mudança sonora (LABOV, 1972; ECKERT; RICKFORD, 2002), é aqui avaliada tendo em vista que novamente Sangster (2002) encontrou variações inter e intra-individuais quanto a diferentes estilos.

### 4. *Liderança*

Propomos a análise da variável *Liderança*, com base na proposta de Labov (2001, v. 2), segundo a qual os líderes da mudança sonora apresentariam índices mais altos de realização da variável inovadora, enquanto que os não-líderes apresentariam índices mais baixos dessa realização. O fato de existirem, em um mesmo grupo, indivíduos avançados (líderes) e conservadores (não-líderes) com relação à mudança sonora aponta para o fato de haver variabilidade no comportamento individual. Por essa razão, a liderança é aqui avaliada.

### 5. *Indivíduo*

Como este estudo objetiva investigar variações inter e intra-individuais, já apontadas em trabalhos diversos (Cf. capítulo 2), propomos a análise da variável *Indivíduo*. Esta deve

nos possibilitar ter uma maior compreensão do caráter das representações lingüísticas, mais especificamente, das representações fonológicas.

Na próxima seção, abordamos as coletas que foram realizadas para o estudo longitudinal, com 2 dos 12 informantes avaliados no estudo transversal tratado acima.

#### **4.5.2 Coletas do estudo longitudinal**

##### **4.5.2.1 Alguns esclarecimentos a respeito das coletas do estudo longitudinal**

Para o estudo longitudinal, foram realizadas, de setembro de 2005 a março de 2006, seis coletas<sup>29</sup> com um intervalo médio de um mês entre uma coleta e outra, com cada um dos 2 participantes selecionados em tal estudo (Cf. participantes L2R e N3Z no QUADRO 6 da seção 4.3)<sup>30</sup>: 1 líder (mulher, jovem, classe média) proveniente do Corpus 2 (FONTES MARTINS, 2001) e 1 não-líder (homem, velho, classe média) proveniente do Corpus 3 (CRISTÓFARO SILVA, 2003).

---

<sup>29</sup> No APÊNDICE D, apresentamos uma breve descrição das coletas realizadas para o estudo longitudinal, bem apresentamos características gerais de cada um dos 2 participantes desse estudo.

<sup>30</sup> Segundo a codificação proposta na seção 4.3 deste capítulo (Cf. QUADRO 6), a informante líder tem o código L2R, e o informante não-líder tem o código N3Z.

Dentre os 12 informantes analisados no estudo transversal, escolhemos esses 2 informantes, por estes terem se mostrado mais acessíveis, mais abertos a colaborar com a presente pesquisa<sup>31</sup>. Ainda, selecionamos 2 participantes por este ser um número habitual de informantes para um estudo longitudinal que, muitas vezes, é feito com base, apenas, em 1 informante (Cf. BONILHA, 2004; CARDOSO, 2005).

As seis coletas mensais, que tiveram uma duração média de 30 minutos cada uma para cada um dos dois informantes, foram feitas com a intenção de se verificar a produção lingüística dos 2 participantes em momentos diferentes no tempo. O número de 6 coletas mensais foi adotado devido à própria disponibilidade de tempo de que dispúnhamos para realizá-las. Também, como será visto no capítulo 6 – no qual é realizada a análise dos dados do estudo longitudinal –, esse número já foi significativo para observarmos variabilidade (ou não) no comportamento desses indivíduos quanto aos fenômenos em análise.

As seis interações com os dois informantes foram feitas buscando-se uma fala o mais natural e espontânea. Dessa forma, as coletas foram realizadas geralmente na casa dos 2 informantes, a fim de se obter uma maior naturalidade na fala desses entrevistados, por este ser um ambiente que lhe é mais íntimo, familiar e até por uma questão de comodidade para eles que preferiram que as coletas fossem assim feitas, em sua

---

<sup>31</sup> Merece nota a nossa experiência com informantes adultos no estudo longitudinal. Deve ser destacada a dificuldade, para o pesquisador, em manter os informantes (no nosso caso, adultos) motivados a colaborar para uma pesquisa de estudo longitudinal com várias interações (no nosso caso, 6 coletas mensais). Como nos dias de hoje, as pessoas, de forma geral, têm muitas ocupações, é difícil encontrar alguém disposto a colaborar com uma pesquisa de coleta de dados. Mais difícil ainda é conseguir que as pessoas que começaram a colaborar para um estudo longitudinal se sintam estimuladas a manter essa colaboração da primeira até a última coleta. Assim, é muito importante ao pesquisador ter paciência, tolerância e sensibilidade para, na hora certa, motivar os seus informantes a cooperar com o seu estudo quando estes parecem desanimados. É, por vezes, inevitável um envolvimento de amizade entre pesquisador e informantes para que estes últimos não desistam de colaborar com a pesquisa, o que pode até inviabilizar a concretização de um estudo longitudinal dependendo da situação. Agradecemos a todos os nossos 12 informantes a participação nesta pesquisa, mas agradecemos em especial aos 2 desses 12 informantes que possibilitaram a realização do estudo longitudinal exatamente como havíamos previsto.

maioria, nas suas próprias casas. Na verdade, com o informante não-líder, todas as 6 coletas foram feitas em sua casa, visto que ele não se mostrou afeito a coletas fora de sua residência. Já com a informante líder, foram realizadas 4 coletas em sua casa e 2 coletas em um restaurante próximo ao seu trabalho, no seu horário de almoço.

As coletas foram baseadas em conversas o mais descontraídas possível. Os temas foram variados: a pesquisadora deixava os informantes conversarem, praticamente, sobre o que quisessem, conduzindo-os a falar sobre o que gostavam, para eles terem mais prazer e, por consequência, mais naturalidade ao conversar. Assim, por exemplo, como o informante não-líder gostava muito de falar sobre a sua família e sobre arte – especialmente sobre música e cinema antigos –, a pesquisadora conduzia a conversa para esses tópicos. Já a informante líder, como gostava de falar sobre o seu trabalho, os seus estudos e os seus amigos, era levada pela pesquisadora a falar desses temas de que gostava. Deve-se ressaltar que, como cada uma das seis coletas com cada um dos dois informantes teve duração média de meia hora, obtivemos um total de, aproximadamente, 6 horas de gravação no estudo longitudinal.

#### **4.5.2.2 Dados**

Como se apontou anteriormente, o estudo longitudinal contou com 6 coletas feitas com 2 dos 12 informantes selecionados nesta pesquisa (Cf. participantes L2R e N3Z no QUADRO 6 da seção 4.3). Nesse estudo, foram considerados todos os dados referentes aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) discutidos anteriormente (Cf. seção 4.4.1). Tais dados são apresentados, analisados e discutidos no capítulo 6.

#### 4.5.2.3 Variáveis analisadas

No estudo longitudinal, também avaliamos a variável dependente *Cancelamento*. Ainda, considerando-se os objetivos, a base teórica do presente trabalho e o caráter próprio do estudo longitudinal, propomos a análise das seguintes variáveis: *Indivíduo*, *Coleta* e *Palavra*<sup>32</sup> (Cf. seção 4.5.1.3.4 para as variáveis *Indivíduo* e *Palavra*).

Objetivamos analisar, em um estudo longitudinal, o comportamento de indivíduos adultos quanto aos fenômenos fonológicos aqui avaliados. Esse tipo de estudo é muito realizado em trabalhos, por exemplo, de aquisição da linguagem com crianças (BONILHA, 2004; MIRANDA, 2007; OLIVEIRA GUIMARÃES, em andamento), mas não o é em trabalhos como o atual que analisa fenômenos de variação sonora em indivíduos adultos.

Ainda, propomos *Coleta* como uma variável, tendo em vista a própria natureza do estudo longitudinal que, por definição, apresenta mais de uma coleta. Considerando Paiva e Duarte (2003), sabemos que é possível encontrar variabilidade individual quando se tem mais de uma coleta com os mesmos indivíduos, o que nos propomos aqui verificar. Já a variável *Palavra* é analisada tendo em vista o que foi dito em 4.5.1.3.4 sobre o fato de considerarmos a proposição do Modelo de Difusão Lexical de que a mudança sonora se difunde gradualmente no léxico. Passemos à próxima seção, que trata do equipamento de gravação que utilizamos nesta pesquisa.

---

<sup>32</sup> No capítulo 6, em que analisamos os dados do estudo longitudinal, também, além de analisarmos essas 3 variáveis individualmente, avaliamos-las em interação, ao verificar a relação da variável *Indivíduo* com as variáveis *Coleta* e *Palavra*.

#### 4.5.3 Equipamento de gravação utilizado

As gravações do estudo transversal e do estudo longitudinal foram feitas utilizando-se um pequeno gravador digital<sup>33</sup> da marca Olympus (VN-240PC – 32 MB, 22 KHz, 8 Bits) no modo *High Quality*. Por ser de tamanho menor e por apresentar uma presilha, tal gravador, na coleta, era posicionado junto ao informante, por exemplo, dentro do bolso de sua camisa. Optamos por esse gravador porque, além de ele proporcionar uma gravação de boa qualidade (devido às suas especificações técnicas citadas acima), é bem pequeno, o que minimiza o desconforto do entrevistado com relação à aparelhagem de gravação.

Ainda, foi utilizado um microfone de lapela da marca Le Son (ML-70/ 70-D). Esse microfone também apresentava presilha e, com isso, era preso à roupa do informante, de modo a também minimizar o desconforto da entrevista causado pela aparelhagem de gravação.

#### 4.5.4 Esclarecimentos sobre a análise estatística realizada

Nesta pesquisa, os dados considerados – dos estudos transversal e longitudinal –, depois de coletados, transcritos, cadastrados e categorizados, foram analisados auditivamente.

---

<sup>33</sup> As dimensões do referido gravador são: 9,9 centímetros de altura por 3,9 de largura e 2,1 cm de profundidade. O peso é de 68 gramas, incluindo as duas pilhas palito utilizadas no gravador.

Após essa análise auditiva, os dados foram submetidos ao programa estatístico Statistical Package for Social Sciences (doravante SPSS) for Windows, v. 13.0<sup>34</sup>.

Neste trabalho, utilizamos o modelo de regressão do programa SPSS tanto na avaliação dos dados do estudo transversal quanto do estudo longitudinal. Esse modelo estatístico permite que se analise “a variabilidade de um fenômeno em relação a um conjunto de fatores” (OLIVEIRA, A., 2005, p. 64), para que se saiba, por exemplo, no nosso caso, por que um determinado indivíduo apresenta pronúncias diferentes para uma mesma palavra.

Utilizamos o modelo de regressão logística binária<sup>35</sup>, porque a variável dependente deste estudo, *Cancelamento*, apresenta apenas duas possibilidades categóricas<sup>36</sup>: cancelamento ou não-cancelamento dos sons em análise. Nas palavras de Oliveira, A. (2005, p. 64), “Esse modelo é utilizado quando se quer investigar, dado um conjunto de possíveis variáveis independentes, quais delas estão de fato associadas à variável dependente”<sup>37</sup>. É preciso dizer que, em nossas análises de regressão logística, trabalhamos com dados de peso relativo. Contudo, como o programa SPSS não apresenta resultados de peso relativo diretamente, tivemos de calculá-lo por meio de tais

---

<sup>34</sup> Uma única análise foi feita utilizando-se o programa estatístico Minitab for Windows, versão 13, para dados do estudo longitudinal, o que será visto na seção 6.2, do capítulo 6.

<sup>35</sup> As análises dos dados dos estudos transversal e longitudinal com regressão logística binária do SPSS são feitas com base em: método *enter* e *categorical, deviation*.

<sup>36</sup> Em consonância com Pierrehumbert (2002), reconhecemos a existência da gradiente fonética; sabemos que há efeitos gradientes nos julgamentos fonotáticos. Contudo, neste trabalho, por questões metodológicas, optamos por analisar os fenômenos em estudo segundo essa perspectiva de duas possibilidades categóricas: cancelamento e não-cancelamento dos sons observados.

<sup>37</sup> Para maiores informações sobre o modelo estatístico “Regressão logística” e até mesmo sobre parâmetros estatísticos diversos (nível de significância, p-valor, probabilidade, razão de chances, chance, desvio da média, etc.) ou sobre o programa estatístico SPSS, consultar Oliveira, A. (2005). Essa obra é uma referência interessante quanto à questão do uso da estatística na lingüística. O autor chega a fazer uma comparação entre os programas Varbrul (amplamente utilizado nos trabalhos feitos em uma perspectiva sociolingüística) e SPSS. Ao fazer uma avaliação de tais programas, o autor aponta que ele optou pelo SPSS. Contudo, para trabalhos em análise variacionista, ele sugere também o programa estatístico R (disponível em: <<http://www.r-project.org/>>. Acesso em ago. 2006), gratuito e de código aberto.

análises, as quais têm como um resultado o valor de “Exp(B)”, que é usado para obter o peso relativo com a seguinte fórmula:  $\text{Exp(B)} / (\text{Exp(B)} + 1) = \text{valor de peso relativo}$ .

Por meio do programa SPSS, também realizamos análises de contingência<sup>38</sup>, as quais, além de apresentarem um panorama geral dos dados, apontam para importantes resultados. Tais análises são apresentadas e discutidas para os dados dos estudos transversal e longitudinal, ao longo dos capítulos 5 e 6, respectivamente.

Um outro tipo de análise do programa SPSS aqui utilizado é o de classificação em árvore pelo método *Chi-squared Automatic Interaction Detector* (doravante, CHAID). O método CHAID permite que se agrupem as categorias de variáveis independentes em grupos homogêneos, chamados *nós*, com relação à variável dependente. Vale mencionar que as árvores de classificação pelo método CHAID, apresentadas nos dois próximos capítulos, são as próprias saídas do programa SPSS.

#### 4.6 Sumário

Neste capítulo, foi apresentada e discutida a metodologia empregada neste estudo. Primeiramente, discutimos o método adotado, o Estudo Descritivo Exploratório, momento no qual apontamos os dois tipos de coleta aqui realizados: 1) uma coleta com todos os 12 indivíduos selecionados, para o estudo transversal e 2) seis coletas com 2

---

<sup>38</sup> Segundo Oliveira, A. (2006, p. 67), a tabela de contingência é “uma tabela de cruzamento de dados de duas variáveis categóricas”. Nas análises de contingência do programa SPSS, é feita uma relação entre a variável dependente e uma variável independente de cada vez, sendo que se pode obter resultado de teste de Chi-square (ou Qui-quadrado). As tabelas de análise de contingência, assim com as de análise de regressão, apresentadas nos dois próximos capítulos, são baseadas nas saídas do programa SPSS. Tais tabelas passaram por uma edição, a fim de serem mostradas neste trabalho.

desses 12 indivíduos, para o estudo longitudinal. Posteriormente, tratamos da amostra trabalhada e do material aqui utilizado, de todo o aparato e estímulos. Por fim, apresentamos e discutimos os procedimentos de coleta e análise dos dados dos estudos transversal e longitudinal.

Nos próximos dois capítulos, analisamos os dados coletados neste estudo. Nessa ocasião, apresentamos a análise estatística realizada, discutindo os resultados obtidos. Como explicitamos neste capítulo, lidamos com dois tipos de banco de dados que refletem dois tipos de estudo diferentes. Propomos uma análise em separado para cada um desses bancos. Isso é feito a seguir, nos capítulos 5 e 6 que analisam, respectivamente, o banco de dados do estudo transversal e o banco de dados do estudo longitudinal.

## CAPÍTULO 5

### ESTUDO TRANSVERSAL: ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

#### 5.1 Introdução

Neste capítulo, abordamos o estudo transversal aqui realizado com os 12 participantes, apresentando e analisando os dados e discutindo os resultados de tal estudo ao longo da seção 5.2. A seção 5.3 sumariza este capítulo.

#### 5.2 Visão geral dos dados do estudo transversal

O banco de dados do estudo transversal conta com um total de 4.955 ocorrências para as 30 palavras dos 3 fenômenos em análise, o que pode ser conferido na TAB. 7:

**TABELA 7: Resultado geral dos dados do estudo transversal**

Cancelamento		Total
<b>não-cancela</b>	n	4093
	%	82,6%
<b>cancela</b>	n	862
	%	17,4%
Total		4955

Como se pode ver pela TAB. 7, do total de 4.955 ocorrências, o cancelamento ocorreu em 862 (17,4%), enquanto que, em 4093 (82,6%), ele não ocorreu. O resultado de 17,4% de cancelamento nos dados do estudo transversal aponta para a existência de variabilidade, mesmo que esse número seja relativamente pequeno. Assim, avaliamos que esse resultado é relevante por até oferecer indícios de representações múltiplas (BYBEE, 2001; PIERREHUMBERT, 2001b), devido à existência de variabilidade.

Aprofundando em nossa análise, conforme mencionado no capítulo anterior, para o estudo transversal, avaliaremos as seguintes variáveis independentes nesta ordem:

- 1) *Fenômeno*;
- 2) *Palavra*;
- 3) *Estilo*;
- 4) *Liderança*;
- 5) *Indivíduo*.

As 5 variáveis serão tratadas separadamente da seção 5.2.1 a 5.2.5. Deve-se destacar que, na análise das variáveis *Liderança* (seção 5.2.4) e *Indivíduo* (seção 5.2.5), avaliamos a interação de cada uma delas com as outras variáveis estudadas: *Fenômeno*, *Palavra* e *Estilo*. Assim, para estas 3 variáveis, fazemos três tipos de análise:

- a) com todos os 12 indivíduos ao mesmo tempo (da seção 5.2.1 a 5.2.3);
- b) com os indivíduos líderes e não-líderes separadamente (seção 5.2.4.1 a 5.2.4.3);
- c) com cada um dos 12 indivíduos em separado (seção 5.2.5.1 a 5.2.4.3).

Além de fazermos a análise (a) para todos os indivíduos simultaneamente, propomos a análise (b) para líderes e não-líderes em separado, visto que estamos considerando, neste estudo, a questão da liderança (LABOV, 2001, v. 2) que, inclusive, serviu como critério para a seleção dos nossos participantes (Cf. seção 4.3). Ainda, por analisarmos o comportamento do indivíduo (OLIVEIRA, M., 1992; SANGSTER, 2002) nesta pesquisa, fazemos a análise (c) para cada indivíduo em separado.

Iniciamos a análise das variáveis de (1) a (5) acima, tratando, na próxima seção, da variável *Fenômeno*. A avaliação desta variável no presente estudo tem como motivação maior o estudo de Sangster (2002), no qual a autora, analisando diferentes fenômenos do inglês, encontrou variabilidade individual relevante.

### 5.2.1 Análise da variável *Fenômeno*

Para iniciar a análise da variável *Fenômeno*, vejamos a TAB. 8:

**TABELA 8: Análise da variável *Fenômeno***

	<b>Fenômenos</b>			Total
	1. Cancelamento de “r” #	2. Cancelamento de líquida V-V	3. Cancelamento de líquida em CC	
<b>Não-Cancelamento</b>	720 80,3%	2575 82,9%	798 83,9%	4093 82,6%
<b>Cancelamento</b>	177 19,7%	532 17,1%	153 16,1%	862 17,4%
<b>Peso relativo</b>	.54	.49	.47	
<b>Total</b>	897	3107	951	4955

Chi-Square Tests: 0,095

A TAB. 8 apresenta os resultados referentes a cada um dos três fenômenos em estudo. Logo abaixo de tal tabela, é apresentado o valor do teste de Chi-square que foi de 0,095. Esse valor indica que não foi estatisticamente significativa a diferença entre as categorias da variável independente em análise, no caso *Fenômeno*, quanto ao número de cancelamento/ não-cancelamento<sup>39</sup>. O fato de o teste de Chi-square não ter apresentado significância em relação à variável *Fenômeno* pode demonstrar que os Fenômenos 1 (r #), 2 (v-v) e 3 (cc) se encontram em um estágio de mudança parecido. Isso porque os percentuais de cancelamento de tais fenômenos são bem próximos (Cf. TAB. 8: os valores de 19,7%, 17,1% e 16,1% de cancelamento), o que foi indicado pelo teste de Chi-square.

Quanto aos resultados de peso relativo presentes na TAB. 8 para os Fenômenos 1 (r #), 2 (v-v) e 3 (cc), notamos que os valores são próximos, assim como ocorreu com os dados percentuais. Contudo, percebemos que o Fenômeno 1 (r #) é o único que favorece o cancelamento com um valor de .54, enquanto que os Fenômenos 2 (v-v) e 3 (cc) desfavorecem com .49 e .47, respectivamente.

Apesar de haver uma pequena diferença de apenas .07 no peso relativo dos dois fenômenos que mais se afastam (os Fenômenos 1 (r #) e 3 (cc)), de modo diverso dos dados percentuais, os resultados de peso relativo da TAB. 8 podem estar indicando que o Fenômeno 1 (r #) encontra-se em um estágio um pouco mais avançado de mudança do que os Fenômenos 2 (v-v) e 3 (cc). De fato, o Fenômeno 1 (r #) parece diferenciar-se dos Fenômenos 2 (v-v) e 3 (cc), tendo em vista que os falantes do português brasileiro

---

<sup>39</sup> O valor do nível de significância obtido pelo teste de Chi-square é arbitrário. Contudo, em lingüística e em outras ciências, especialmente da área de humanas, o valor utilizado para medir o nível de significância é 0,05 (abaixo desse valor, existe significância e acima, não). Deve-se salientar que, como na TAB. 8, logo abaixo de algumas tabelas deste trabalho, apresentamos o valor do teste de Chi-square, quando os resultados deste forem relevantes.

têm alguma consciência dele (LABOV, 2001, v. 2)<sup>40</sup>, não tendo consciência dos outros dois (FONTES MARTINS, 2001; CRISTÓFARO SILVA, 2003). Contudo, seria interessante que trabalhos futuros investigassem mais detidamente essa questão.

Nas seções 5.2.4.1 e 5.2.5.1, aprofundaremos a análise da variável *Fenômeno*, ao abordar a relação desta com as variáveis *Liderança* e *Indivíduo*, respectivamente. Como neste estudo consideramos que a mudança sonora se difunde gradualmente no léxico que é plástico e dinâmico (CHEN; WANG, 1975; BYBEE, 2001; PIERREHUMBERT, 2001b), refletimos que tão importante quanto avaliar o comportamento dos fenômenos é avaliar o comportamento das palavras referentes a eles. Isto é o que fazemos na próxima seção, em que abordamos a variável *Palavra*.

### **5.2.2 Análise da variável *Palavra***

Para examinar a variável *Palavra*, vejamos a TAB. 9 que apresenta uma análise com o percentual de cancelamento e o peso relativo de cada uma das 30 palavras referentes aos 3 fenômenos aqui estudados:

---

<sup>40</sup> Quando se quer imitar um “caipira”, é comum cancelar, intencionalmente, o “r” final em palavras, como *mulher*, *colher* e *amor* (*muié*, *cuié*, *amô*). Isso seria uma evidência para alguma consciência que as pessoas têm do Fenômeno 1 (r #) em análise.

**TABELA 9: Análise da variável *Palavra***

Palavras	Cancelamento				Percentual médio de cancelamento	Total	PR
	não-cancela		cancela				
	n	%	n	%			
mulher	91	68,9%	41	31,1%	19,7%	132	.87
jogador	47	74,6%	16	25,4%		63	.83
maior	60	76,9%	18	23,1%		78	.81
qualquer	75	78,9%	20	21,1%		95	.80
lugar	86	79,6%	22	20,4%		108	.79
dólar	57	80,3%	14	19,7%		71	.78
melhor	100	84,7%	18	15,3%		118	.72
governador	47	85,5%	8	14,5%		55	.71
computador	66	85,7%	11	14,3%		77	.71
mar	91	91,0%	9	9,0%		100	.59
eles	169	71,0%	69	29,0%	17,1%	238	.86
para	691	75,4%	225	24,6%		916	.83
óculos	60	81,1%	14	18,9%		74	.77
ele	660	81,5%	150	18,5%		810	.77
agora	193	86,9%	29	13,1%		222	.69
dinheiro	143	90,5%	5	9,5%		158	.60
ela	291	93,3%	21	6,7%		312	.51
era	199	96,1%	8	3,9%		207	.37
feira	82	98,8%	1	1,2%		83	.15
Lula	87	100%	0	0%		87	.00
outro	77	60,2%	51	39,8%	16,1%	128	.91
sempre	84	71,2%	34	28,8%		118	.86
quatro	49	77,8%	14	22,2%		63	.81
problema	98	78,4%	27	21,6%		125	.80
livro	64	79,0%	17	21,0%		81	.79
primeiro	75	93,8%	5	6,3%		80	.49
três	106	98,1%	2	1,9%		108	.22
presidente	73	98,6%	1	1,4%		74	.17
grande	83	98,8%	1	1,2%		84	.15
Brasil	89	98,9%	1	1,1%		90	.14
<b>Total</b>	<b>4093</b>	<b>82,6%</b>	<b>862</b>	<b>17,4%</b>		<b>4955</b>	

Na TAB. 9, as palavras relacionadas ao Fenômeno 1 (r #) estão destacadas de cinza claro; as palavras relacionadas ao Fenômeno 2 (v-v) não apresentam qualquer destaque; as palavras relacionadas ao Fenômeno 3 (cc) estão destacadas de cinza escuro. Ainda, observando as palavras de cada fenômeno individualmente, nota-se que elas se encontram em ordem decrescente quanto ao percentual de cancelamento (5ª coluna da TAB. 9).

A TAB. 9 nos mostra que, apesar de o cancelamento médio para as 30 palavras ser de 17,4%, os índices de cancelamento para cada palavra individualmente variam de 0% a 39,8%, assim como o peso relativo varia de .00 a .91.

Assim, pela TAB. 9, percebemos que as 30 palavras apresentam percentuais de cancelamento bem diversos, mesmo quando uma palavra de um fenômeno – por exemplo, a palavra *Brasil* do Fenômeno 3 (cc), que apresenta 1,1% de cancelamento – é comparada com outra palavra desse mesmo fenômeno – por exemplo, com a palavra *outro* também do Fenômeno 3 (cc), que, inclusive, é a palavra que apresenta o maior percentual de cancelamento (39,8%). Note-se que, em *outro*, o som cancelado – a líquida [r] – encontra-se em sílaba átona postônica (*outro*) que, como se apontou no capítulo 4, é o contexto sonoro que mais favorece a aplicação do Fenômeno 3 (cc) (CRISTÓFARO SILVA, 2003). Nesse resultado relativo a *outro*, vemos que a palavra é importante na implementação da mudança sonora (CHEN; WANG, 1975; BYBEE, 2001), mas que o fator estrutural também é relevante, por poder estar atuando como um estabilizador desta (OLIVEIRA, M., 1992).

Vale salientar que apenas as 10 palavras do Fenômeno 1 (r #) apresentam peso relativo superior a .50, ou seja, todas elas favorecem o cancelamento. Quanto ao Fenômeno 2 (v-v), 7 das 10 palavras favorecem e, em relação ao Fenômeno 3 (cc), 5 das 10 palavras favorecem. O próprio fato de o percentual médio de cancelamento das palavras do Fenômeno 1 (r #) – 19,7% – ser maior do que os das palavras dos Fenômenos 2 (v-v) – 17,1% – e 3 (cc) – 16,1% – reflete isso de certa maneira (Cf. 6ª coluna da TAB. 9). Assim, esse resultado está de acordo com a análise feita para a variável *Fenômeno* na seção anterior, 5.2.1, em que se levantou a hipótese de o Fenômeno 1 (r #) encontrar-se em um estágio de mudança um pouco mais avançado do que os Fenômenos 2 (v-v) e 3

(cc). É possível que o Fenômeno 1 (r #) esteja lexicalmente mais difundido (CHEN; WANG, 1975; BYBEE, 2001), por ter atingido mais palavras do que os outros dois fenômenos.

Convém destacar que, em todo esse corpus do estudo transversal, a palavra que apresenta 0% de cancelamento é *Lula*. Isso pode estar ligado ao fato de que *Lula* é um nome próprio. Oliveira, M. (1991) aponta que os nomes próprios atuam como inibidores da mudança sonora. Inclusive, este pode ser o caso da palavra supracitada *Brasil* que, como *Lula*, é um nome próprio, e apresentou um percentual muito baixo de cancelamento (1,1%).

No entanto, é interessante notar que a palavra *Brasil*, mesmo que pouco, ao menos chega a apresentar cancelamento, o que não ocorre com *Lula*. A nossa hipótese para tal fato é que *Lula* apresenta um contexto sonoro que dificulta mais o cancelamento sonoro do que *Brasil*: *Lula* apresenta duas líquidas idênticas [l] em, relativamente, pouco material fonético (quatro sons), ou seja, apresenta uma líquida em cada uma de suas duas sílabas de dois sons. As líquidas são conhecidas por serem sons de articulação mais complexa, sendo, por esse motivo, apontadas como os últimos sons, no português, a serem adquiridos pelas crianças em fase de aquisição (YAVAS, 1988; HERNANDORENA, 1990; LAMPRECHT, 1993)<sup>41</sup>. Vemos aqui também, no caso da palavra *Lula*, a importância de se considerar o fator estrutural na análise da mudança sonora (OLIVEIRA, M., 1992).

---

<sup>41</sup> De acordo com Balen (1997, p. 26): “É de consenso dos autores que estudam a aquisição fonológica e fonética do Português, tais como YAVAS (1988) e HERNANDORENA (1990), que as líquidas são os últimos fones a serem adquiridos pela criança em seu desenvolvimento lingüístico, bem como, de acordo com HERNANDORENA (1988), LAMPRECHT (1993) e BALEN (1995), as líquidas são os segmentos que mais comumente aparecem alterados em crianças e adultos.” Esta última afirmação de Balen, a de que as líquidas são freqüentemente modificadas por crianças e adultos, é um fato que se observa em nossa língua com relação aos diversos fenômenos de mudança sonora que as envolvem, como os Fenômenos 1 (r #), 2 (v-v) e 3 (cc) avaliados neste trabalho.

Quanto ao peso relativo das palavras em análise, como se pode perceber pela TAB. 9, também mesmo palavras relacionadas a um único fenômeno apresentam valores diferentes de peso relativo. Como exemplo, podemos citar as palavras *sempre* (que, com .86, favorece o cancelamento) e *grande* (que, com .15, desfavorece o cancelamento), ambas relacionadas ao Fenômeno 3 (cc).

O fato de *sempre* favorecer o fenômeno de cancelamento, possivelmente, se deve a uma questão de o som cancelado em tal palavra estar em sílaba átona postônica (*sempre*). Ao contrário, a palavra *grande* desfavorece por o som cancelado nela estar em sílaba tônica (*grande*), a qual tende a inibir fenômenos de cancelamento sonoro (FONTES MARTINS, 2001; HUBACK, 2003; CRISTÓFARO SILVA, 2003). Novamente, com a análise dos itens *sempre* e *grande*, notamos a importância de se refletir sobre o fator estrutural (OLIVEIRA, M., 1992), conjugado com o léxico (BYBEE, 2001).

Essa discrepância nos percentuais e pesos relativos das palavras em estudo (TAB. 9) passa agora a ser avaliada por meio de uma análise de classificação em árvore (método CHAID). Isso porque, nessa análise em árvore, as 30 palavras são divididas em diferentes grupos (nós) de acordo com a homogeneidade, em termos de percentual de cancelamento, que existe entre elas<sup>42</sup>. Dessa forma, é possível perceber, por tal análise, como as 30 palavras se relacionam, se agrupam quanto à semelhança nos índices de cancelamento delas. Isso é o que vemos na FIG. 6:

---

<sup>42</sup> Maiores detalhes sobre as análises de classificação em árvore se encontram na seção 4.5.4 do capítulo 4, “Metodologia”.

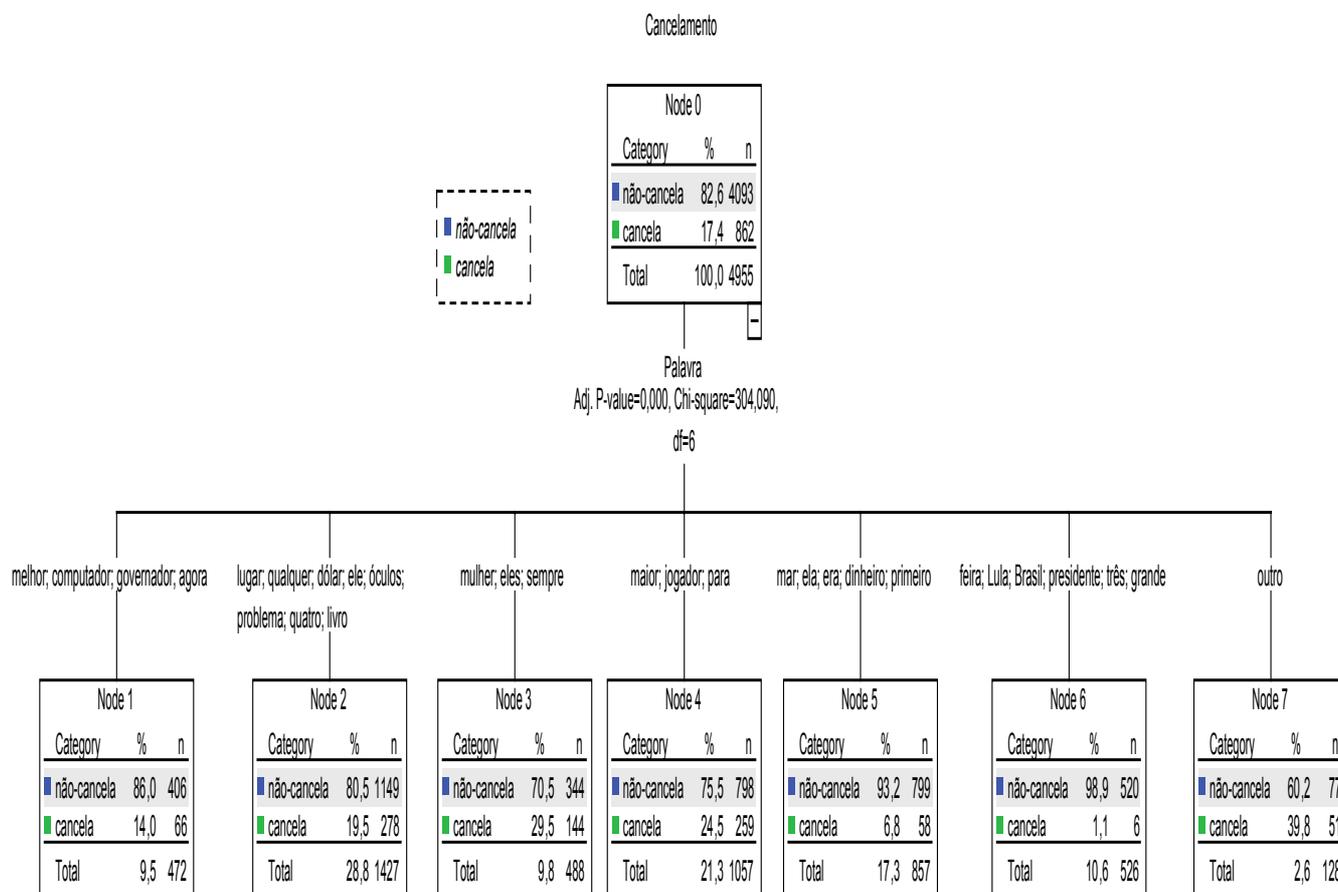


FIGURA 6 – Classificação em árvore (método CHAID) para as 30 palavras do estudo transversal

Pela análise da árvore da FIG. 6, podemos fazer uma primeira observação: as 30 palavras em estudo apresentam uma diferença que é estatisticamente significativa, já que o programa SPSS agrupou-as em 7 nós/ grupos. Pode-se perceber que os percentuais médios de cancelamento são bem distintos para os diferentes nós/ grupos de palavras: o nó com menor percentual (6º nó) apresenta uma média de 1,1% de cancelamento, enquanto que o nó com maior percentual (7º nó) apresenta 39,8% de cancelamento. Nessa análise em árvore, em consonância com os resultados discutidos para a TAB. 9,

também vemos que mesmo palavras que se relacionam a um único fenômeno se comportam de modo diferente. Esse é o caso, por exemplo, das palavras *eles* (3º nó) e *Lula* (6º nó). Ambas essas palavras se relacionam ao Fenômeno 2 (v-v), no entanto, elas se encontram em nós com percentuais médios de cancelamento bem distintos: enquanto *eles* está no 3º nó que apresenta 29,5% de cancelamento, *Lula* está no 6º nó que apresenta 1,1% (Cf. FIG. 6).

A análise da FIG. 6 aponta para a dinamicidade e a plasticidade do léxico (OLIVEIRA, M., 1995<sup>43</sup>; BYBEE, 2001; PIERREHUMBERT, 2001b) que apresenta um caráter probabilístico (BOD; HAY; JANNEDY, 2003), na medida em que palavras relacionadas a um mesmo fenômeno não têm um comportamento homogêneo e categórico, como seria esperado em uma proposta de caráter neogramático (OSTHOFF; BRUGMAN, 1878), segundo a qual a mudança sonora é lexicalmente abrupta. Assim, essa discrepância no comportamento de palavras de um mesmo fenômeno aponta para padrões de difusão lexical da mudança sonora (BYBEE, 2001; CHEN; WANG, 1975).

Pela FIG. 6, nota-se que as palavras que estão no grupo das que menos cancelam (com uma média de 1,1% de cancelamento) são: *feira*, *Lula*, *Brasil*, *presidente*, *três*, *grande* (6º nó na FIG. 6). Como foi apontado ainda nesta seção, é interessante notar que *Lula* é a única das 30 palavras que não apresenta nenhum caso de cancelamento, e que as outras palavras desse grupo apresentam um baixo percentual de cancelamento que vai de 1,1% a 1,9%, com 1 ou 2 ocorrências de cancelamento em números absolutos (Cf. TAB. 9). Além disso, *Lula*, assim como *Brasil*, é um nome próprio, o qual, como vimos, é um inibidor de fenômenos de mudança sonora (OLIVEIRA, M., 1991). Vale destacar que *presidente*, por ser uma palavra formal (OLIVEIRA, M., 1995), também inibe tais

---

<sup>43</sup> Como Oliveira, M. (1995, p. 87) propõe que o fator frequência seja atribuído a cada item na situação de fala em que ele ocorre – considerando-se, inclusive, os falantes individuais –, tal autor aponta para a plasticidade e dinamicidade do léxico.

fenômenos. Ainda, *três* e *grande* têm o som a ser cancelado em sílaba tônica, o que desfavorece o cancelamento sonoro (CRISTÓFARO SILVA, 2003), assim como o fato de *três* ser monossílabo desfavorece. Como o monossílabo apresenta naturalmente pouco material fonético, este tende a ser preservado, a fim de que a palavra monossilábica não seja descaracterizada.

Por outro lado, como também foi mencionado nesta seção, a palavra *outro* foi a que apresentou o maior percentual de cancelamento (39,8%). A representação em árvore da FIG. 6 nos mostra algo interessante em relação ao item *outro*, algo que não foi possível ser visualizado pela análise da TAB. 9. Pela FIG. 6, notamos que a palavra *outro* se difere de todas as demais palavras, já que se encontra sozinha em um nó (o 7º nó), não sendo agrupada com nenhuma outra palavra. Isso indica que a diferença entre o percentual de cancelamento de *outro* e de todas as demais palavras é estatisticamente significativa. Tal achado, mais uma vez, aponta para padrões de difusão lexical, já que a análise aqui realizada sinaliza para o fato de que observar o comportamento de palavras individualmente é importante (CHEN; WANG, 1975; BYBEE, 2001). Seria interessante que trabalhos futuros investigassem se não existem outras propriedades específicas aos itens *Lula* e *outro*, por exemplo, com relação a efeitos de frequência (BYBEE, 2001)<sup>44</sup>. Ainda neste capítulo, a variável *Palavra* será retomada quando da análise de sua relação com as variáveis *Liderança* (seção 5.2.4.2) e *Indivíduo* (seção 5.2.5.2).

Neste trabalho, adotamos alguns aspectos metodológicos da Sociolinguística (LABOV, 1966), por estarmos lidando com fenômenos de variação sonora. Dentre tais aspectos, consideramos a variável *Estilo*, sendo que, especialmente pelo fato de Sangster (2002)

---

<sup>44</sup> Como destacamos no capítulo 4, “Metodologia”, neste estudo não pudemos considerar efeitos de frequência (BYBEE, 2001), tendo em vista que só trabalhamos com palavras mais frequentes, não avaliando palavras menos frequentes.

ter encontrado variabilidade individual relevante com relação a diferentes estilos, propomo-nos a avaliar tal variável, o que fazemos na próxima seção.

### 5.2.3 Análise da variável *Estilo*

Para avaliar a variável *Estilo*, vejamos a TAB. 10 que apresenta uma análise com o percentual de cancelamento e o peso relativo dos 3 estilos aqui avaliados:

**TABELA 10: Análise da variável *Estilo***

	Estilos			Total
	Fala espontânea	Nomeação por figura	Leitura	
<b>Não-Cancelamento</b>	3102 80,0%	319 89,4%	672 93,3%	4093 82,6%
<b>Cancelamento</b>	776 20,0%	38 10,6%	48 6,7%	862 17,4%
<b>Peso relativo</b>	.71	.43	.35	
<b>Total</b>	3878	357	720	4955

Chi-square: 0,000

Pela análise da TAB. 10, notamos que, para os 3 fenômenos em análise, os percentuais de cancelamento em cada um dos três estilos são um tanto distantes uns dos outros, sendo que o estilo *fala espontânea* apresenta o maior percentual (20%), seguido da *tarefa de nomeação por figura* (10,6%) e da *tarefa de leitura* (6,7%). Inclusive, o valor do teste de Chi-square de 0,000 demonstra que é estatisticamente significativa tal diferença entre os três estilos. Esse resultado está de acordo com a teoria sociolinguística (LABOV, 1972), segundo a qual, a ocorrência de fenômenos de mudança sonora varia de acordo com os estilos citados.

Ainda na TAB. 10, o resultado do peso relativo para a variável *Estilo* também aponta para esse aspecto da sociolinguística. O estilo *fala espontânea* é o único que favorece os fenômenos de cancelamento com .71. Já os estilos *tarefa de nomeação por figura* com .43 e *tarefa de leitura* com .35 desfavorecem o cancelamento, sendo que este último desfavorece mais do que o primeiro. O maior desfavorecimento da *leitura* se deve ao fato de esta tarefa ser ainda mais formal do que a *nomeação por figura*.

Podemos afirmar que, na TAB. 10, os resultados percentuais e de peso relativo para o *Estilo* estão em consonância com o que os estudos sobre mudança e variação sonora (LABOV, 1972; CHAMBERS, 1995; SANGSTER, 2002) afirmam a respeito de tal variável: falantes usam diferentes estilos com propriedades linguísticas distintas (ABAURRE, 1981). Diferenças estilísticas requerem do falante habilidades cognitivas diferenciadas na articulação da linguagem (BYBEE, 2001).

Sabe-se que, em situação de *fala espontânea*, do ponto de vista linguístico (LABOV, 1972), o falante fica mais à vontade e, com isso, tende a falar com mais naturalidade, o que favorece a ocorrência de fenômenos fonológicos, como a lenição. Ao contrário, em situação de *tarefa de nomeação por figura* e, mais ainda, em situação de *tarefa de leitura*, o falante policia a sua fala, de modo que a ocorrência de fenômenos fonológicos é significativamente menor do que na *fala espontânea*.

Uma possível consequência dessa diferença entre tais estilos são variações na taxa de elocução (*speech rate*, Cf. BARBOSA, 2000)<sup>45</sup>. Como o falante fica mais à vontade em

---

<sup>45</sup> Barbosa (2000) utiliza o termo *taxa de elocução* por considerá-lo mais adequado para traduzir *speech rate*. O autor afirma que taxa de elocução é um termo “mais apropriado que “velocidade de fala”, pois a grandeza física geralmente usada para indicá-la, “sílabas por unidade de tempo”, não expressa a velocidade real de deslocamento dos articuladores da fala (como a mandíbula), que é obtida pelo uso de técnicas modernas de registro do movimento de produção de fala (como eletromiografia, microfones de raios X, entre outras)” (BARBOSA, 2000, p. 402).

situação de *fala espontânea*, é provável que a taxa de elocução nesse estilo seja consideravelmente maior do que no estilo *tarefa de nomeação por figura* e, mais ainda, do que no estilo *tarefa de leitura*. Ainda, é possível que esse “relaxamento” do falante na *fala espontânea* leve a uma redução da magnitude dos gestos articulatórios<sup>46</sup> (BROWMAN; GOLDSTEIN, 1992; ALBANO, 2001), o que não ocorreria na *nomeação de figura* nem na *leitura*. Seria interessante que trabalhos futuros investigassem essa relação entre estilo e taxa de elocução, assim como entre estilo e gesto articulatório.

Tendo em vista os resultados desta seção, avaliamos que, quanto à variável *Estilo*, atua o caráter probabilístico da linguagem (BOD; HAY; JANNEDY, 2003). Há uma gradiência (PIERREHUMBERT, 2002) entre os diferentes estilos. Por exemplo, como vimos, *nomeação de figura* e *leitura* são formais, entretanto, não na mesma medida, já que esta é ainda mais formal do que aquela. Para representar a gradualidade que há, em nossos dados, entre os estilos informal e formal, apresentamos a FIG. 7:

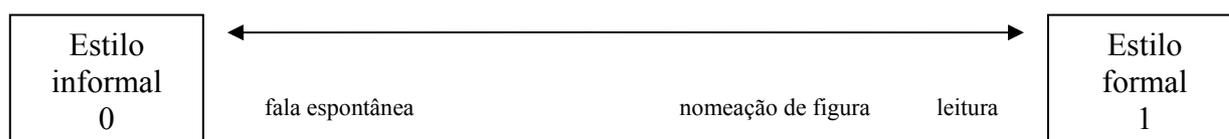


FIGURA 7 – Categorias estilo informal e estilo formal em um contínuo de 0 (estilo informal) a 1 (estilo formal)

Avaliando a FIG. 7, vemos que as categorias estilo informal (0) e estilo formal (1) estão situadas, cada uma, em um dos extremos da reta que representa o contínuo gradual entre essas duas categorias (PIERREHUMBERT, 2002). À esquerda, do lado do estilo

<sup>46</sup> De acordo com Albano (2001, p. 54): “O gesto não corresponde ao segmento ou ao traço, ficando a meio caminho entre os dois. Um [b], por exemplo, envolve apenas a coordenação de um gesto de oclusão labial com um gesto de adução das pregas vocais, enquanto um [m] envolve os mesmos gestos coordenados ao gesto vélico de abertura, cujo início e fim não precisam coincidir inteiramente com os daqueles.”

informal, encontra-se a *fala espontânea*. Já à direita, do lado do estilo formal, encontram-se a *nomeação de figura* e a *leitura*. Contudo, a *leitura* está projetada mais à direita do que a *nomeação de figura*, o que indica a maior formalidade daquela em relação a esta. Assim, por meio da FIG. 7, confirmamos que a noção de gradualidade (de 0 a 1) está presente em nossos dados (PIERREHUMBERT, 2002). Tal noção aponta para o aspecto dinâmico da linguagem, o que a noção de categoricidade absoluta (0 ou 1) não o faz.

A fim de visualizar de modo mais claro os resultados apresentados nesta seção e até mesmo de aprofundar a nossa análise, abordamos a variável *Estilo* por meio da classificação em árvore (método CHAID), com todos os três fenômenos aqui avaliados ao mesmo tempo e, posteriormente, com um desses fenômenos de cada vez. Assim, nesta primeira análise, a classificação em árvore presente na FIG. 8 confirma que, com relação aos 3 fenômenos ao mesmo tempo, há uma diferença estatisticamente significativa entre os 3 estilos aqui considerados, visto que estes se encontram em nós separados em tal árvore:

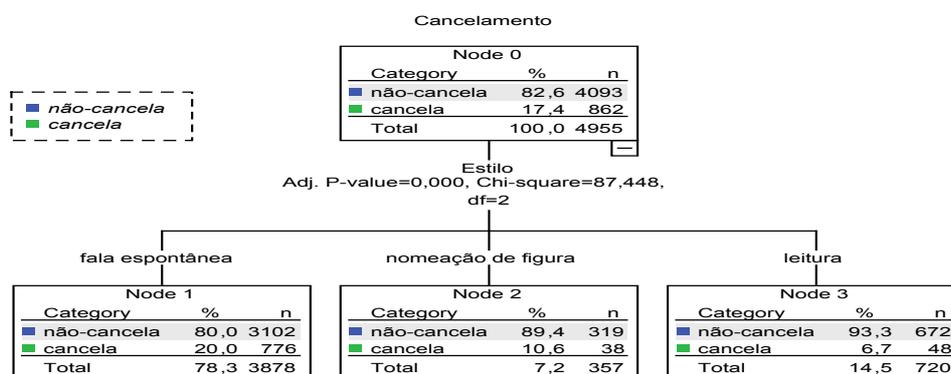


FIGURA 8 – Classificação em árvore (método CHAID) para a variável *Estilo*, em relação aos três fenômenos em análise ao mesmo tempo

Contudo, quando fazemos uma análise dos 3 estilos com cada um desses 3 fenômenos em separado, obtemos um resultado um tanto diferente. Vejamos, assim, a FIG. 9 que apresenta uma classificação em árvore para o Fenômeno 1 (r #) apenas:

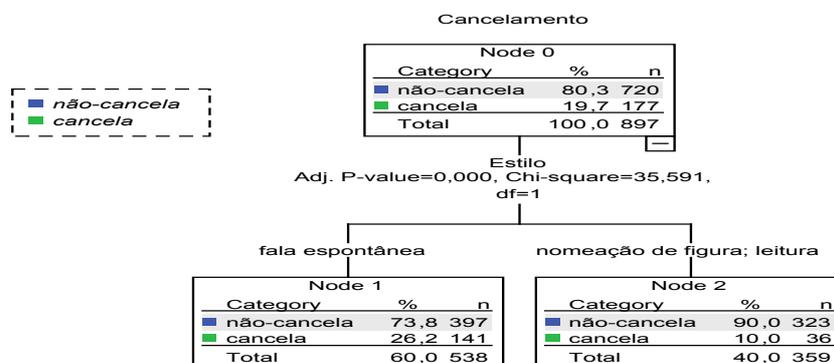


FIGURA 9 – Classificação em árvore (método CHAID) para a variável *Estilo*, em relação ao Fenômeno 1 (r #)

Avaliando a árvore da FIG. 9 que relaciona a variável *Estilo* ao Fenômeno 1 (r #) somente, percebemos uma diferença com relação à análise feita para todos os 3 fenômenos ao mesmo tempo (Cf. FIG. 8). Nesta análise, como vimos, os 3 estilos ficaram em nós separados, o que demonstra que a diferença entre eles é estatisticamente significativa. Já na análise da árvore que considera apenas o Fenômeno 1 (r #) (Cf. FIG. 9), os estilos foram agrupados em dois nós, sendo que os estilos *nomeação de figura* e *leitura* foram agrupados em um mesmo nó e a *fala espontânea*, em um outro nó diferente. Isto demonstra que, nessa última análise, não há diferença estatisticamente significativa entre os estilos *nomeação de figura* e *leitura*. Só há diferença entre esses dois estilos juntos e o estilo *fala espontânea*. Vejamos, agora, a FIG. 10 que apresenta uma classificação em árvore para o Fenômeno 2 (v-v) apenas:

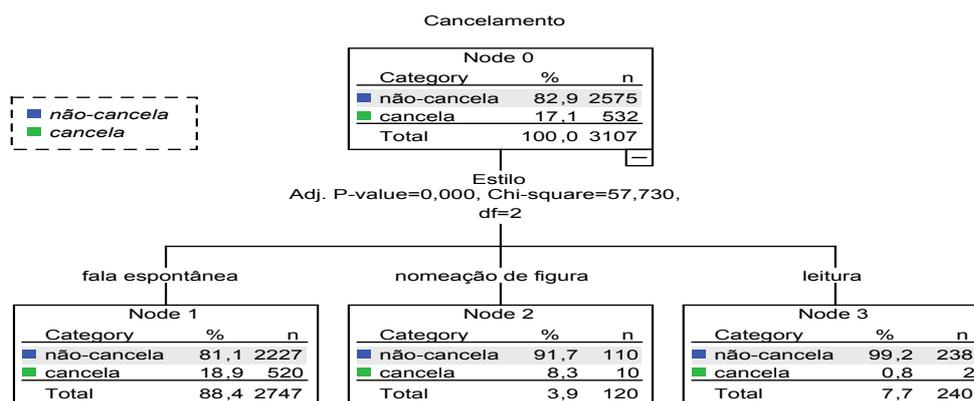


FIGURA 10 – Classificação em árvore (método CHAID) para a variável *Estilo*, em relação ao Fenômeno 2 (v-v)

Na árvore da FIG. 10 que considera apenas o Fenômeno 2 (v-v) – assim como na árvore da FIG. 8 que considera todos os 3 fenômenos ao mesmo tempo –, a diferença entre os 3 estilos foi estatisticamente significativa, visto que cada um destes ficou em um nó separado. Isso não ocorreu na análise da FIG. 11 que considera apenas o Fenômeno 3 (cc):

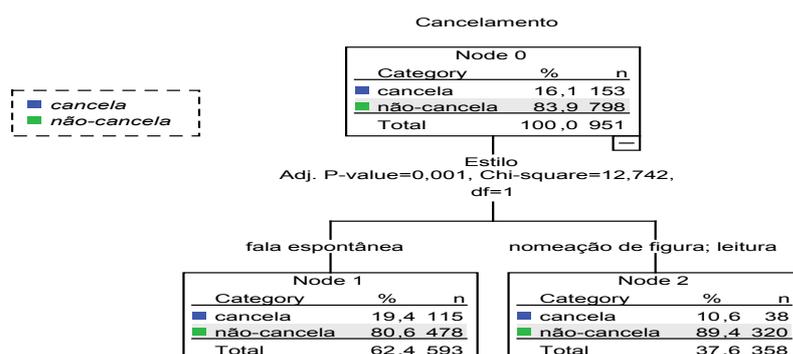


FIGURA 11 – Classificação em árvore (método CHAID) para a variável *Estilo*, em relação ao Fenômeno 3 (cc)

Assim como na análise do Fenômeno 1 (r #) apenas (FIG. 9), essa análise com o Fenômeno 3 (cc) (FIG. 11) mostra que *nomeação de figura* e *leitura* não apresentam diferença estatisticamente significativa, já que em que esses dois estilos foram agrupados

em um único nó. Somente a diferença da *nomeação de figura* e *leitura* juntas em relação à *fala espontânea* é que é significativa.

Os resultados encontrados nas análises da FIG. 9 a FIG. 11 – feitas para os 3 estilos em relação aos 3 diferentes fenômenos separadamente – indicam que há variação no comportamento dos indivíduos quanto ao estilo e que, interessantemente, este pode apresentar resultados diferentes de acordo com o fenômeno. Assim, tipos de estilo podem ser estatisticamente diferentes quanto a um fenômeno (Fenômeno 2 (v-v), Cf. FIG. 10), mas não quanto a um outro (Fenômeno 1 (r #), Cf. *nomeação de figura* e *leitura* na FIG. 9).

Vale notar que, nas análises dos três fenômenos (Cf. FIG. 9 a 11), o que vemos é que o estilo *fala espontânea*, que é informal, se mostrou diferente estatisticamente dos outros dois estilos, *nomeação de figura* e *leitura*, que são formais. Inclusive os resultados de peso relativo na TAB. 10 para os três estilos apontam para esse fato de que a *fala espontânea* se encontra distante (sendo o único estilo que favorece os fenômenos de cancelamento em estudo) dos outros dois estilos que se encontram próximos (ambos desfavorecendo tais fenômenos). Esses achados corroboram o que a literatura em mudança e variação sonora afirma a respeito de haver diferenças entre estilos informais e formais (LABOV, 1972). O estilo *fala espontânea* parece funcionar como um “motor” da variabilidade, por ser informal justamente, favorecendo fenômenos de mudança sonora (ABAURRE, 1981; LABOV, 1972; SANGSTER, 2002).

O *Estilo* será aqui retomado na análise de sua relação com a *Liderança* (seção 5.2.4.3) e o *Indivíduo* (seção 5.2.5.3). Na próxima seção, abordamos a variável *Liderança* (LABOV, 2001, v. 2). Propomos a análise dessa variável, tendo em vista que o fato de

existirem, em um mesmo grupo, indivíduos avançados (líderes) e conservadores (não-líderes) com relação à mudança sonora aponta para a variabilidade no comportamento individual.

#### 5.2.4 Análise da variável *Liderança*

Vejamos a TAB. 11 que compara os percentuais de cancelamento dos líderes e dos não-líderes, apresentando os valores de peso relativo de tais indivíduos:

**TABELA 11: Análise da variável *Liderança***

	<b>Liderança</b>		<b>Total</b>
	<b>Líderes</b>	<b>Não-líderes</b>	
<b>Não-Cancelamento</b>	1800 70,4%	2293 95,7%	4093 82,6%
<b>Cancelamento</b>	758 29,6%	104 4,3%	862 17,4%
<b>Peso Relativo</b>	.75	.25	
<b>Total</b>	2558	2397	4955

A TAB. 11 mostra que, considerando-se todos os 3 fenômenos ao mesmo tempo, a diferença nos percentuais de cancelamento dos líderes e dos não-líderes é grande: enquanto todos os líderes juntos apresentam um percentual de 29,6% de cancelamento, todos os não-líderes juntos apresentam apenas 4,3%, o que é quase 7 vezes menor que o percentual dos líderes. Vale lembrar que o percentual médio de cancelamento dos dados do estudo transversal é de 17,4% (Cf. TAB. 7). Com isso, vemos que os líderes apresentam um percentual médio de cancelamento consideravelmente superior ao do estudo transversal (quase o dobro), enquanto que os não-líderes apresentam um percentual médio de cancelamento muito inferior ao de tal estudo (quase 4 vezes

menor). Esse fato indica que existe um comportamento de tendência, segundo o qual, em conjunto, os líderes se comportam como líderes, assim como, também em conjunto, os não-líderes se comportam como não-líderes (LABOV, 2001, v. 2).

Ainda, como se pode notar pela TAB. 11, a diferença de peso relativo entre os indivíduos líderes e os não-líderes é grande. Enquanto os líderes favorecem os fenômenos de cancelamento em análise com um alto valor de peso relativo de .75, os não-líderes desfavorecem tais fenômenos com um peso relativo baixo de .25. Esses resultados de peso relativo também nos mostram que tanto os líderes quanto os não-líderes se comportam como esperado (LABOV, 2001, v. 2), com os líderes tendendo a utilizar mais a variável inovadora, ao contrário dos não-líderes que tendem a utilizá-la pouco.

A seguir, da seção 5.2.4.1 a 5.2.4.3, abordamos a relação da variável *Liderança* com as variáveis *Fenômeno*, *Palavra* e *Estilo*, respectivamente. O objetivo dessa abordagem é aprofundar nossa análise verificando se líderes e não-líderes se comportam de forma esperada, de acordo com a sua respectiva categoria quanto a fenômenos, palavras e estilos diversos. Vale lembrar que fazemos o mesmo à frente para a variável *Indivíduo* (da seção 5.2.5.1 a 5.2.5.3).

#### **5.2.4.1 Relação entre as variáveis *Liderança* e *Fenômeno***

Nesta seção, abordamos a relação da variável *Liderança* com a variável *Fenômeno*, avaliando os três fenômenos estudados em conjunto e separadamente. Em uma primeira

análise, tratamos desses fenômenos e dos indivíduos líderes e não-líderes em separado. Assim, a TAB. 12 apresenta os percentuais de cancelamento dos indivíduos líderes apenas, para cada um dos 3 fenômenos em estudo:

**TABELA 12: Relação das variáveis *Fenômeno e Liderança*: somente líderes**

	Fenômenos			Total
	1. Cancelamento de "r" final em nominais	2. Cancelamento de líquida V-V	3. Cancelamento de líquida em CCV	
<b>Não-Cancelamento</b>	330 69,6%	1124 69,5%	346 74,2%	1800 70,4%
<b>Cancelamento</b>	144 30,4%	494 30,5%	120 25,8%	758 29,6%
<b>Total</b>	474	1618	466	2558

Chi-Square Tests: 0,127

Pela análise da TAB. 12, percebe-se que os líderes apresentam um percentual de cancelamento maior quanto aos Fenômenos 1 (r#) (30,4%) e 2 (v-v) (30,5%) do que quanto ao Fenômeno 3 (cc) (25,8%). Vale destacar que tais percentuais são quase o dobro do que o percentual médio de cancelamento dos dados do estudo transversal (17,4%, Cf. TAB. 7). Esse fato novamente indica que os líderes desta pesquisa, quando analisados em conjunto, se comportam como líderes (LABOV, 2001, v. 2), tendo em vista que inclusive apresentam um índice de realização de cada um dos fenômenos em estudo superior à média geral dos dados para todos os fenômenos em conjunto.

Nota-se, ainda, que a diferença entre os três fenômenos para os líderes é estatisticamente insignificante, o que é demonstrado pelo valor de Chi-square abaixo da TAB. 12 (0,127). Esse fato sugere que os líderes em conjunto avaliam tais fenômenos de modo homogêneo (SANGSTER, 2002).

Passemos agora à análise dos indivíduos não-líderes, para observar se estes vão seguir o mesmo padrão dos líderes ou não. Desse modo, a TAB. 13 apresenta o percentual de cancelamento apenas dos indivíduos não-líderes para cada um dos 3 fenômenos em estudo:

**TABELA 13: Relação das variáveis *Fenômeno* e *Liderança*: somente não-líderes**

	Fenômenos			Total
	1. Cancelamento de "r" final em nominais	2. Cancelamento de líquida V-V	3. Cancelamento de líquida em CCV	
<b>Não-Cancelamento</b>	390 92,2%	1451 97,4%	452 93,2%	2293 95,7%
<b>Cancelamento</b>	33 7,8%	38 2,6%	33 6,8%	104 4,3%
<b>Total</b>	423	1489	485	2397

Chi-Square Tests: 0,000

Como se pode perceber pela análise da TAB. 13, os não-líderes apresentam um percentual de cancelamento maior quanto ao Fenômeno 1 (r #) (com 7,8%), sendo que o Fenômeno 3 (cc) (com 6,8%) vem em segundo lugar, e o Fenômeno 2 (v-v) (com 2,6%) vem em terceiro. Ainda, nota-se que esses percentuais são muito inferiores ao percentual médio de cancelamento dos dados do estudo transversal (17,4%, Cf. TAB. 7). Esse fato também indica que os não-líderes desta pesquisa, em conjunto, se comportam como não-líderes (LABOV, 2001, v. 2).

Pelo valor do Chi-square (0,000) abaixo da TAB. 13, percebemos que a diferença de cancelamento entre os Fenômenos 1 (r #), 2 (v-v) e 3 (cc), para os não-líderes, é estatisticamente significativa. Esse resultado sugere que, ao contrário dos líderes, os não-líderes em conjunto avaliam tais fenômenos de modo heterogêneo, ou seja, podem apresentar um percentual de cancelamento para um fenômeno diverso do de outro (SANGSTER, 2002). A fim de verificar como essa avaliação dos não-líderes ocorre especificamente, fazemos uma análise de classificação em árvore na FIG. 12:

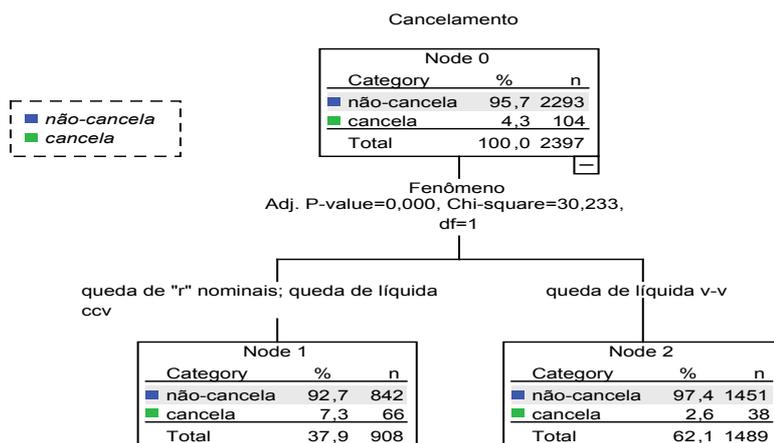


FIGURA 12 – Classificação em árvore (método CHAID) para os indivíduos não-líderes somente, quanto aos três fenômenos em análise

Pela análise da FIG. 12, vemos que, para os indivíduos não-líderes, há diferença estatisticamente significativa dos Fenômenos 1 (r #) e 3 (cc) em relação ao Fenômeno 2 (v-v). Como os Fenômenos 1 (r #) e 3 (cc) foram agrupados em um mesmo nó (1º nó da FIG. 12) nessa análise de árvore, isso indica que tais fenômenos apresentam um comportamento semelhante, mais homogêneo. Pelo fato de o Fenômeno 2 (v-v) estar situado em um nó sozinho (2º nó da FIG. 12), isso mostra que este fenômeno, sim, se comporta de forma diferente em relação aos Fenômenos 1 (r #) e 3 (cc). Pelo fato de esse resultado relativo aos não-líderes ter sido diferente do resultado dos líderes, vemos que líderes e não-líderes podem ter padrões diferentes com relação a tais fenômenos (BYBEE, 2001, SANGSTER, 2002): enquanto os líderes tratam-nos de um modo homogêneo, os não-líderes tratam-nos, em parte, de modo heterogêneo.

A fim de avaliar se líderes e não-líderes apresentam comportamentos diferentes quanto às palavras em estudo, abordamos, na próxima seção, a relação da variável *Liderança* com a variável *Palavra*.

### 5.2.4.2 Relação entre as variáveis *Liderança* e *Palavra*

Para tratar da relação entre as variáveis *Liderança* e *Palavra*, propomos uma análise em separado para líderes e não-líderes. Assim, comparemos as árvores das FIG. 13 e 14 que mostram como as palavras em estudo são distribuídas para líderes e para não-líderes, respectivamente.

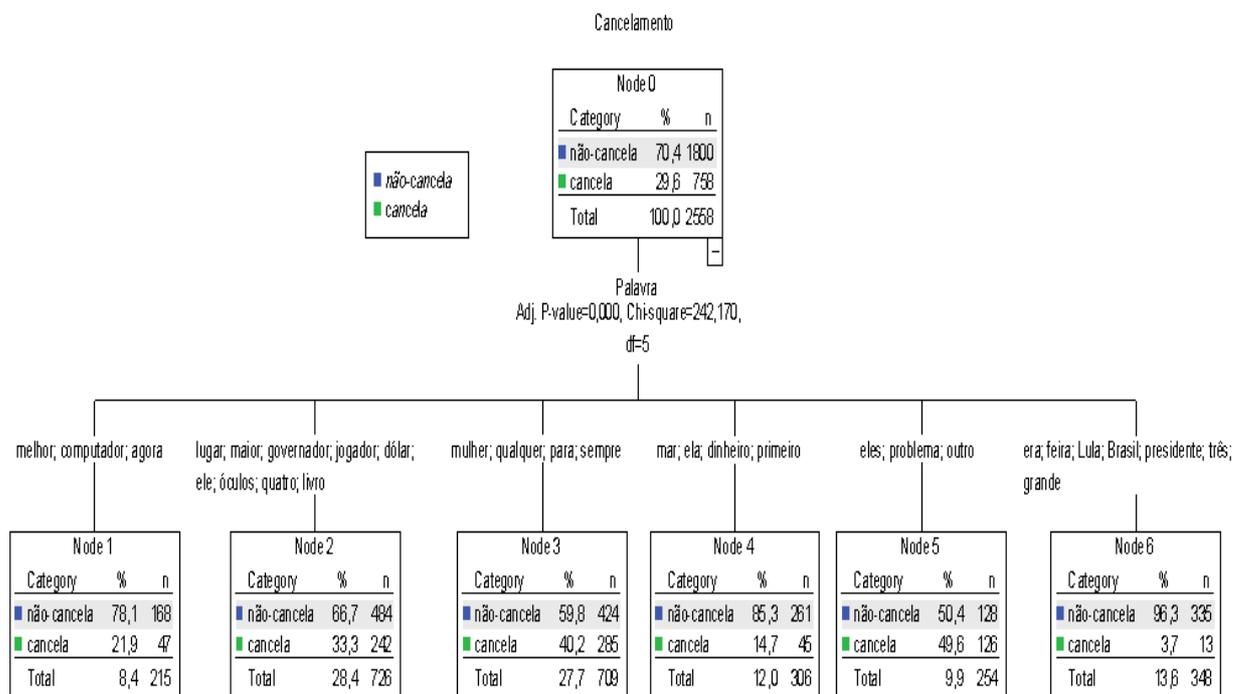


FIGURA 13 – Classificação em árvore (método CHAID) com as 30 palavras do estudo transversal somente para os indivíduos líderes

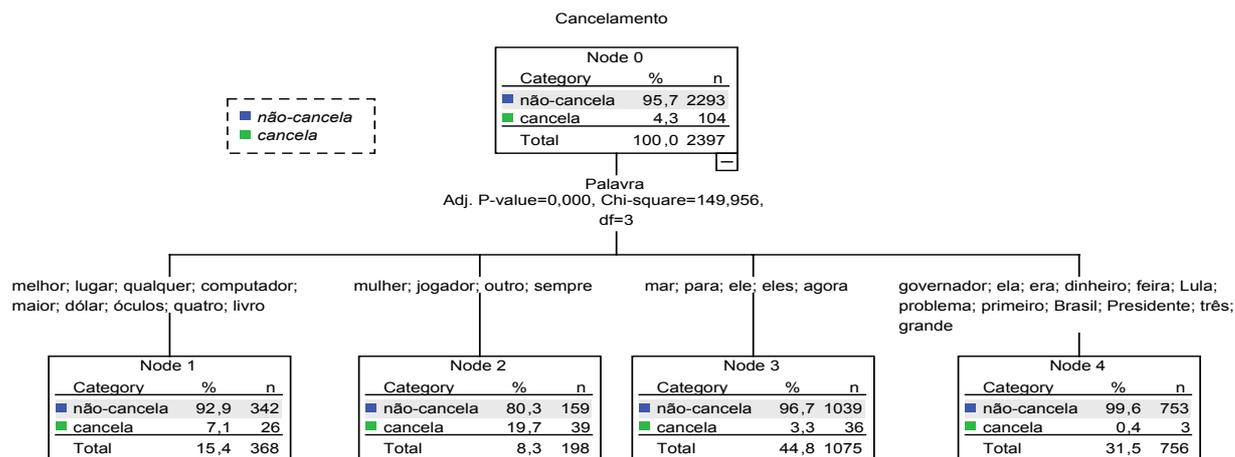


FIGURA 14 – Classificação em árvore (método CHAID) com as 30 palavras do estudo transversal somente para os indivíduos não-líderes

Pela análise das FIG. 13 e 14, percebemos que indivíduos líderes e não-líderes agrupam as 30 palavras em estudo de forma diferente: enquanto os líderes agrupam-nas em 6 nós (FIG. 13), os não-líderes agrupam-nas em 4 nós (FIG. 14). Podemos, assim, fazer uma primeira observação ao comparar os agrupamentos das palavras feitos pelos líderes e pelos não-líderes. A dispersão das 30 palavras em mais nós para os indivíduos líderes mostra que há uma heterogeneidade maior no comportamento dessas palavras em relação a estes indivíduos, ou seja, tal dispersão mostra que é provável que os líderes apresentam, no uso, maior número de exemplares em competição (PIERREHUMBERT, 2001b) e, até mesmo, maior multiplicidade de exemplares (BYBEE, 2001; PIERREHUMBERT, 2001b) do que os não-líderes.

Mais interessante ainda a ser observado na comparação dos agrupamentos das FIG. 13 e 14 acima é que, em geral, os indivíduos líderes e não-líderes agrupam as palavras que apresentam menor percentual de cancelamento de maneira parecida. Contudo, a forma como eles agrupam as palavras que apresentam maior percentual de cancelamento é um tanto diferente. Esse fato também pode ter relação com efeitos de frequência (BYBEE,

2001). Palavras mais freqüentes apresentariam maior variabilidade ou multiplicidade de exemplares do que palavras menos freqüentes, que apresentam menor número de exemplares e, conseqüentemente, menor multiplicidade destes (BYBEE, 2001; PIERREHUMBERT, 2001b). Contudo, nesse caso, seriam necessários maiores estudos, já que aqui não pudemos nos deter na questão da freqüência, tendo em vista que não tivemos um controle adequado dela na seleção de nossas palavras.

Analisando a FIG. 13 referente aos indivíduos líderes, percebemos que as palavras que apresentam menor percentual de cancelamento são: *era, feira, Lula, Brasil, Presidente, três, grande* (com uma média de 3,7% de cancelamento – ver 6º nó). Já para os indivíduos não-líderes, analisando a FIG. 14, percebemos que as palavras que apresentam menor percentual são: *governador, ela, era, dinheiro, feira, Lula, problema, primeiro, Brasil, Presidente, três, grande* (com uma média de 0,4% de cancelamento – ver 4º nó). Como se pode notar, as palavras que menos cancelam para os líderes estão incluídas nas palavras que menos cancelam para os não-líderes.

Passemos às palavras que apresentam maior percentual de cancelamento para líderes e não-líderes. Vemos pela FIG. 13 que, quanto aos líderes, as palavras que apresentam maior percentual de cancelamento são: *eles, problema, outro* (com uma média de 49,6% de cancelamento – ver 5º nó). Já quanto aos não-líderes, vemos pela FIG. 14 que as palavras que apresentam maior percentual são: *mulher, jogador, outro, sempre* (com uma média de 19,7% de cancelamento – ver 2º nó). Com se pode notar, nessa análise, somente a palavra *outro* coincide para líderes e não-líderes. Mais uma vez<sup>47</sup>, vemos que a palavra *outro*, que se relaciona ao Fenômeno 3 (cc), está mais avançada em termos da mudança sonora que a atinge (CHEN; WANG, 1975; BYBEE, 2001). Isso,

---

<sup>47</sup> Este resultado é compatível com a análise feita na seção 5.2.2, em que *outro*, dentre todas as 30 palavras em estudo, apresentou o maior índice de cancelamento (Cf. TAB. 9).

provavelmente, está ligado a efeitos de frequência (BYBEE, 2001), já que, dentre todas as 10 palavras relacionadas ao Fenômeno 3 (cc), *outro* é a mais freqüente em termos de ocorrência no nosso corpus do estudo transversal, com 128 ocorrências (Cf. TAB. 9). Seria interessante que trabalhos futuros aprofundassem nessa questão da freqüência, tendo em vista que, como ressaltamos, no presente estudo não pudemos fazê-lo, por não termos tido um controle de freqüência na seleção de nossas palavras.

É interessante ressaltar que a palavra *problema* que está dentre as que apresentam maior percentual de cancelamento para os líderes (ver 5º nó da FIG. 13), inversamente, está dentre as que apresentam menor percentual para os não-líderes (ver 4º nó da FIG. 14). A palavra *problema* apresenta uma estrutura que é menos freqüente em termos de tipo no português (BYBEE, 2001; PROJETO ASPA) – dois encontros consonantais tautossilábicos seguidos. Contudo, em termos de freqüência de ocorrência, ela foi a segunda mais freqüente no nosso corpus (com 125 ocorrências, Cf. TAB. 9) dentre as 10 palavras analisadas quanto ao Fenômeno 3 (cc). Sugerimos que tal palavra, por essa questão de freqüência, pode ter maior multiplicidade de exemplares (PIERREHUMBERT, 2001b) para os líderes do que para os não-líderes. Na verdade, por todas as análises realizadas até o momento para líderes e não-líderes, acabamos chegando a esta conclusão: líderes apresentam maior variabilidade e multiplicidade de exemplares de uma mesma palavra (BYBEE, 2001; PIERREHUMBERT, 2001b) do que não-líderes, o que podemos afirmar para o módulo da produção ao menos.

Podemos destacar, ainda, que palavras, como *eles* e *para*, as quais apresentam um baixo percentual médio de cancelamento para os indivíduos não-líderes (ver 3º nó da FIG. 14), apresentam um alto percentual médio de cancelamento para os líderes (ver 3º e 5º nós da FIG. 13). Novamente, a análise feita no parágrafo anterior cabe aqui neste caso das

palavras *eles* e *para*, já que estas palavras – especialmente, *para* – estão dentre as que apresentam maior frequência de ocorrência no nosso corpus, quanto ao Fenômeno 2 (v-v) (Cf. TAB. 9, *para* com 916 ocorrências e *eles* com 238) e, sendo assim, os indivíduos líderes apresentariam maior multiplicidade e variabilidade de exemplares (BYBEE, 2001; PIERREHUMBERT, 2001b) para tais palavras do que os não-líderes.

Na próxima seção, aprofundamos um pouco mais nesta questão, ao analisarmos a variável *Estilo*, para os líderes e os não-líderes em separado.

#### **5.2.4.3 Relação entre as variáveis *Liderança* e *Estilo***

Passemos agora a avaliar a variável *Liderança* com relação ao *Estilo* para líderes e não-líderes em separado. Na árvore da FIG. 15, é feita uma análise em que se relacionam tais variáveis para os indivíduos não-líderes somente:

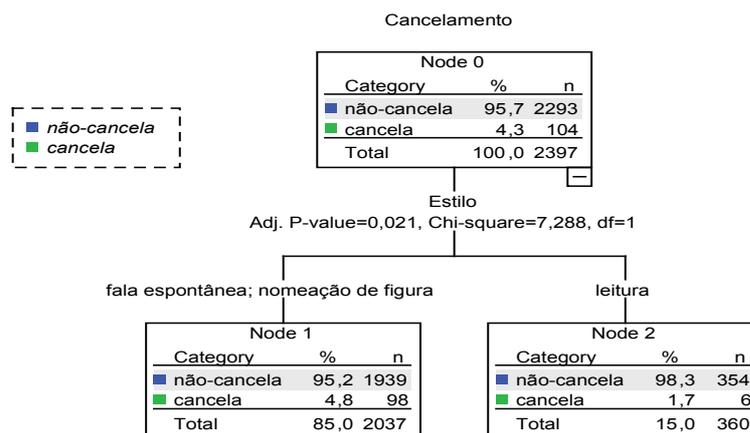


FIGURA 15 – Classificação em árvore (método CHAID) com a relação entre as variáveis *Liderança* e *Estilo* para os não-líderes

Pela análise da árvore da FIG. 15, notamos que, diferentemente do que ocorreu nas análises da seção 5.2.3 (para todos os 3 fenômenos ao mesmo tempo – FIG. 8 – e em separado – FIG. 9, 10 e 11), pela primeira vez, os estilos *fala espontânea* e *nomeação de figura* se encontram agrupados em um mesmo nó, separados do estilo *leitura* que se encontra em outro nó. Assim, por essa análise, vemos que *fala espontânea* e *nomeação de figura* não apresentam diferença estatisticamente significativa para os indivíduos não-líderes. Para esses indivíduos, só há diferença da *fala espontânea* e da *nomeação de figura* com relação ao estilo *leitura*, o que vemos pelo P-valor = 0,021 presente na FIG. 15.

Vejamos, agora, a árvore na FIG. 16 que mostra a relação da variável *Estilo* com os líderes somente:

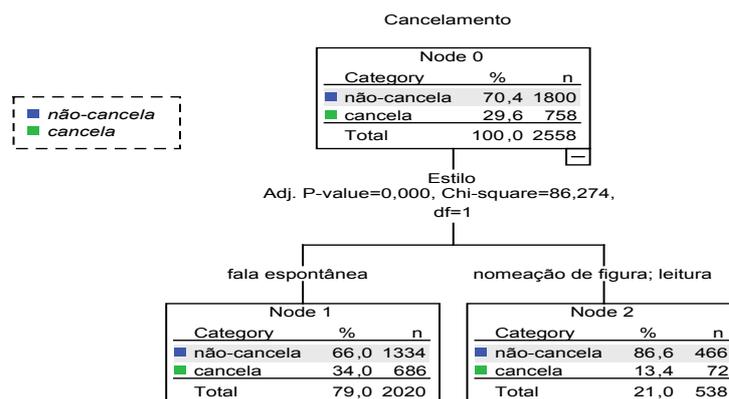


FIGURA 16 – Classificação em árvore (método CHAID) com a relação entre as variáveis *Liderança* e *Estilo* para os líderes

Pela análise da FIG. 16, vemos que os líderes agrupam os 3 estilos de forma diferente dos não-líderes (FIG. 15): os líderes agrupam *nomeação de figura e leitura* em um mesmo nó e *fala espontânea* em um nó diferente (Cf. FIG. 16). Inclusive, o P-valor = 0,000 presente na árvore da FIG. 16 demonstra que, para os indivíduos líderes, essa diferença da *nomeação de figura e leitura* quanto à *fala espontânea* é ainda mais significativa estatisticamente do que a diferença, para os indivíduos não-líderes, da *fala espontânea e nomeação de figura* quanto à *leitura* (ver P-valor = 0,021 da árvore da FIG. 15).

Podemos afirmar que os informantes líderes apresentam um comportamento parecido nos estilos *nomeação de figura e leitura*, apresentando um comportamento diferente no estilo *fala espontânea*. Já os informantes não-líderes apresentam um comportamento parecido nos estilos *fala espontânea e nomeação de figura*, apresentando um comportamento diferente no estilo *leitura*.

Vemos que, enquanto para os líderes há uma distinção bem nítida entre estilo informal (*fala espontânea*) e estilo formal (*leitura e nomeação de figura*) – Cf. FIG. 16 –, para os

não-líderes não o há, já que um estilo, em princípio, informal (*fala espontânea*) foi agrupado com um estilo formal (*nomeação de figura*) na análise da FIG. 15 (LABOV, 1972; ABAURRE, 1981; SANGSTER, 2002). O que parece estar ocorrendo é que, para os não-líderes, a *fala espontânea* é um estilo mais formal do que informal (LABOV, 1972; MARCHUSCHI, 2003). Isso porque, no estilo *fala espontânea*, o percentual de cancelamento apresentado pelos não-líderes (4,8%, Cf. FIG. 16) além de ser muito inferior – praticamente 7 vezes menor – ao percentual de cancelamento apresentado pelos líderes (34%, Cf. FIG. 16), é quase 4 vezes menor do que o percentual de cancelamento de todos os indivíduos líderes e não-líderes juntos (20%, Cf. TAB. 10). Ainda, esse percentual de cancelamento dos não-líderes na *fala espontânea* é quase 3 vezes menor do que o percentual de cancelamento dos líderes nos estilos formais (13,4%, Cf. FIG. 16).

Com base na relação entre as variáveis *Liderança* e *estilo* discutida nesta seção e considerando a FIG. 7 (Cf. seção 5.2.3), vemos que enquanto os indivíduos líderes apresentam bordas bem definidas para os estilos informal (*fala espontânea*) e formal (*nomeação de figura* e *leitura*), os indivíduos não-líderes apresentam bordas menos definidas, com um centro mais complexo para tais estilos. Vejamos a FIG. 17 que busca representar essa diferença entre líderes e não-líderes quanto à variável *Estilo*:

## LÍDERES



## NÃO-LÍDERES

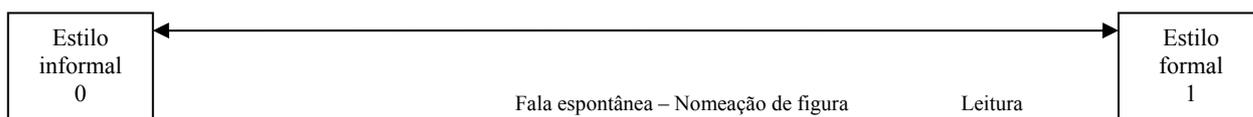


FIGURA 17 – Líderes e não-líderes quanto aos estilos informal (0) e formal (1)

Pela representação da FIG. 17, vemos que para os líderes (1ª reta), o estilo *fala espontânea* está situado à esquerda da reta, do lado do estilo informal (0). Já para os não-líderes (2ª reta), a *fala espontânea* está situada mais no centro, contudo, um tanto projetada à direita, em direção a (1) – estilo formal (PIERREHUMBERT, 2002). Assim, por meio da FIG. 17, vemos que, em relação ao *Estilo*, líderes e não-líderes têm representações diferentes (BYBEE, 2001; PIERREHUMBERT, 2001b).

Na próxima seção, em que passamos a tratar da última variável em estudo, *Indivíduo*, a qual é foco deste trabalho, de certa forma, retomaremos a discussão sobre líderes e não-líderes. Por exemplo, investigaremos se, mais uma vez, individualmente eles apresentam comportamento incompatível com a sua respectiva categoria. A variável *Indivíduo*, assim como a *Liderança*, será avaliada também em interação com as variáveis *Fenômeno*, *Palavra* e *Estilo* (da seção 5.2.5.1 a 5.2.5.3). Esperamos, com a análise do *Indivíduo*, ter uma maior compreensão das representações lingüísticas, mais especificamente, das representações fonológicas.

### 5.2.5 Análise da variável *Indivíduo*

Para iniciar a análise da variável *Indivíduo*, vejamos a TAB. 14 que apresenta os dados percentuais de cancelamento dos 3 fenômenos aqui considerados, para cada um dos 12 indivíduos em estudo<sup>48</sup>:

**TABELA 14: Análise da variável *Indivíduo*: dados percentuais**

Indivíduos	Cancelamento				Total
	Não-cancelamento		Cancelamento		
	n	%	n	%	
Indivíduo 1 – L1W	275	57,1%	207	42,9%	482
Indivíduo 2 – L1V	284	76,5%	87	23,5%	371
Indivíduo 3 – N1A	293	93,3%	21	6,7%	314
Indivíduo 4 – N1L	587	93,8%	39	6,2%	626
Indivíduo 5 – L2L	249	61,3%	157	38,7%	406
Indivíduo 6 – L2R	180	80,0%	45	20,0%	225
Indivíduo 7 – N2L	408	98,3%	7	1,7%	415
Indivíduo 8 – N2M	379	98,2%	7	1,8%	386
Indivíduo 9 – L3R	599	80,2%	148	19,8%	747
Indivíduo 10 – L3J	213	65,1%	114	34,9%	327
Indivíduo 11 – N3Z	318	99,1%	3	0,9%	321
Indivíduo 12 – N3G	308	91,9%	27	8,1%	335
<b>Total</b>	<b>4093</b>	<b>82,6%</b>	<b>862</b>	<b>17,4%</b>	<b>4955</b>

<sup>48</sup> Como esclarecemos no QUADRO 6 da seção 4.3, a codificação proposta para nomear cada um dos 12 participantes desta pesquisa apresenta 3 caracteres. O primeiro caractere são as letras “L” e “N”, que designam *indivíduo líder* e *indivíduo não-líder*, respectivamente. O segundo caractere são os números “1”, “2” e “3” que designam *Corpus 1*, *Corpus 2* e *Corpus 3*, respectivamente. O terceiro e último caractere é a inicial do nome do informante. Assim, o código “L1W” do Indivíduo 1 da TAB. 14 indica que esse informante é líder (“L”), é proveniente do Corpus 1 (“1”) e tem a letra “W” como inicial de seu nome. Vale destacar ainda que a ordem em que os 12 indivíduos aparecem na TAB. 14 é a ordem em que eles foram categorizados no programa SPSS, quando da análise dos dados.

Na TAB. 14, os indivíduos estão agrupados segundo o corpus de que fazem parte. Assim, como é possível perceber pela codificação, pertencem: a) ao Corpus 1, os indivíduos de 1 a 4; b) ao Corpus 2, os indivíduos de 5 a 8; e c) ao Corpus 3, os indivíduos de 9 a 12. É possível perceber pela codificação ainda que os indivíduos líderes têm a inicial “L”, enquanto que os não-líderes têm a inicial “N”. Para melhor visualização, destacamos os líderes em cinza, enquanto que os não-líderes não apresentam destaque.

Pela análise da TAB. 14, percebemos que há uma grande diferença entre os indivíduos líderes e não-líderes quanto ao percentual de cancelamento: enquanto o percentual de cancelamento dos líderes está na casa das dezenas, o dos não-líderes está na casa das unidades (Cf. penúltima coluna da TAB. 14). Esse fato evidencia, mais uma vez (Cf. seção 5.2.4), que, considerando-se os percentuais totais de cancelamento e não-cancelamento de todos os três fenômenos em análise ao mesmo tempo, líderes se comportam como líderes e não-líderes se comportam como não-líderes (LABOV, 2001, v. 2).

Contudo, ainda analisando a TAB. 14, percebemos que há uma considerável diferença no percentual de cancelamento de não-líderes entre si e, especialmente, de líderes entre si. Por exemplo, em relação aos líderes, enquanto o Indivíduo 1 – L1W apresenta 42,9% de cancelamento, o Indivíduo 2 – L1V apresenta 23,5%, ou seja, L1V apresenta quase metade do percentual de L1W. Já quanto aos não-líderes, enquanto o Indivíduo 11 – N3Z apresenta 0,9% de cancelamento, o Indivíduo 12 – N3G apresenta 8,1%, ou seja, N3G apresenta um percentual nove vezes maior do que N3Z. Esse fato aponta para a importância de se analisar o comportamento individual, já que pode haver diferença significativa dos líderes entre si e dos não líderes entre si, o que é trabalhado na próxima

análise que considera dados de peso relativo<sup>49</sup>. Assim, vejamos a TAB. 15 que apresenta o peso relativo de cada um dos 12 indivíduos, considerando todos os 3 fenômenos em estudo ao mesmo tempo:

**TABELA 15: Resultados de peso relativo para aos 12 indivíduos quanto aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em conjunto**

<b>Indivíduos</b>	<b>Peso relativo</b>
Indivíduo 1 – L1W	.94
Indivíduo 2 – L1V	.57
Indivíduo 3 – N1A	.53
Indivíduo 4 – N1L	.50
Indivíduo 5 – L2L	.70
Indivíduo 6 – L2R	.60
Indivíduo 7 – N2L	.15
Indivíduo 8 – N2M	.11
Indivíduo 9 – L3R	.49
Indivíduo 10 – L3J	.85
Indivíduo 11 – N3Z	.13
Indivíduo 12 – N3G	.45

Avaliando a TAB. 15, notamos que os resultados de peso relativo nos mostram que os indivíduos que mais favoreceram os fenômenos de cancelamento em análise foram, em ordem crescente:

- a) Indivíduo 6 – L2R com .60;
- b) Indivíduo 5 – L2L com .70;
- c) Indivíduo 10 – L3J com .85;
- d) Indivíduo 1 – L1W com .94.

<sup>49</sup> Também na próxima seção que trata da relação entre as variáveis *Indivíduo* e *Fenômeno*, tal fato será discutido em maiores detalhes. Nessa ocasião, verificaremos o que ocorre quando fazemos essa análise com os 3 fenômenos em separado.

Os quatro indivíduos que mais favoreceram os fenômenos de cancelamento (a-d) são todos líderes, o que se percebe pela letra inicial “L” da codificação atribuída a cada indivíduo. Esse resultado é compatível com a noção de líder, segundo a qual é líder um indivíduo com maiores índices de realização de uma variável inovadora que reflete uma mudança sonora (LABOV, 2001, v. 2). Alguns indivíduos ficaram em uma faixa intermediária, com peso relativo próximo ou igual a .50, com uns tendendo a favorecer os fenômenos de cancelamento observados:

- e) Indivíduo 4 – N1L com .50;
- f) Indivíduo 3 – N1A com .53;
- g) Indivíduo 2 – L1V com .57;

e outros tendendo a desfavorecer tais fenômenos:

- h) Indivíduo 12 – N3G com .45;
- i) Indivíduo 9 – L3R com .49.

Nessa faixa intermediária, note-se que há líderes e não-líderes. Há 2 não-líderes (e, f) e 1 líder (g) tendendo a favorecer os fenômenos aqui considerados, e 1 não-líder (h) e 1 líder (i) tendendo a desfavorecer. A existência dessa faixa intermediária que apresenta líderes e não-líderes problematiza a noção de líder proposta por Labov (2001, v. 2). Nessa faixa, contrariamente à noção de líder discutida na seção 2.3.1, vemos que há um líder que tende a desfavorecer os fenômenos de cancelamento em estudo (Indivíduo 9 – L3R com .49), assim como há dois não-líderes que tendem a favorecer tais fenômenos (Indivíduo 3 – N1A com .53; Indivíduo 4 – N1L .50). Isso não é esperado dentro dessa

noção. Seria esperado que somente líderes favorecessem esses fenômenos de cancelamento, assim como, ao contrário, somente não-líderes os desfavorecessem.

Há, ainda, os indivíduos que, sem dúvida, desfavorecem os fenômenos de cancelamento em análise. Esses indivíduos são, em ordem crescente:

j) Indivíduo 8 – N2M com .11;

l) Indivíduo 11 – N3Z com .13;

m) Indivíduo 7 – N2L com .15.

É possível perceber que os indivíduos que realmente desfavorecem os fenômenos de cancelamento (j-m) são todos não-líderes, o que é esperado dentro da concepção de liderança de Labov (2001, v. 2), já que, nesta, os não-líderes apresentariam menores índices de realização de uma variável inovadora que reflete uma mudança sonora. Vejamos agora a TAB. 16<sup>50</sup> que permite uma melhor visualização das informações discutidas anteriormente a respeito de todos os indivíduos – líderes e não-líderes – analisados nesta pesquisa:

---

<sup>50</sup> Vale ressaltar que a TAB. 16 será retomada por diversas vezes no decorrer deste capítulo.

**TABELA 16: Classificação dos 12 indivíduos em faixas de peso relativo e quanto ao contínuo gradual não-líder (0) – líder (1)**

LIDERANÇA				
←		→		
0 não-líder		1 líder		
Faixas de peso relativo				
Desfavorecem mais:		Intermediário – tendendo a desfavorecer:	Intermediário – tendendo a favorecer:	Favorecem mais:
de .11 a .15		de .45 a .49	de .50 a .57	de .60 a .94
<b>Indivíduos</b>	N2L com .11	N3G com .45	N1L com .50	L2R com .60
	N3Z com .13	L3R com .49	N1A com .53	L2L com .70
	N2L com .15		L1V com .57	L3J com .85
				L1W com .94

Na TAB. 16, os indivíduos não-líderes estão destacados de cinza claro, enquanto que os indivíduos líderes estão destacados em cinza escuro. Essa tabela permite visualizar com facilidade que os indivíduos que apresentam maior probabilidade de realizar os fenômenos em análise, por terem maior valor de peso relativo, são todos líderes (ver 5ª coluna da TAB. 16). Já os que apresentam menor probabilidade, por terem menor valor de peso relativo, são todos não-líderes (ver 2ª coluna TAB. 16).

Nas terceira e quarta colunas da TAB. 16, os indivíduos que apresentam uma probabilidade intermediária (com um valor de peso relativo mediano, próximo de .50) são líderes e não-líderes. O que vemos em tal tabela é que as suas bordas apresentam não-líderes (2ª coluna) e líderes (5ª coluna) separadamente, mas no seu meio (3ª e 4ª colunas), há uma mistura dessas categorias: poderíamos assim dizer que, na TAB. 16, temos 4 indivíduos líderes, 3 indivíduos não-líderes e 5 indivíduos em um campo intermediário entre não-líder e líder. Com base nesse achado, na TAB. 16, propomos uma reta (semelhante à da FIG. 7 relacionada ao *Estilo*) que se relaciona diretamente com as faixas de peso relativo que vêm logo abaixo dela. Tal reta representa a

gradiência (PIERREHUMBERT, 2002) que há entre as categorias não-líder (0) e líder (1).

A TAB. 16 nos oferece indícios de que a liderança, segundo critérios probabilísticos (BOD; HAY; JANNEDY, 2003), reflete um contínuo gradual (PIERREHUMBERT, 2002) com bordas definidas (2ª e 5ª colunas) e um centro mais complexo (3ª e 4ª colunas). Ou seja, a liderança é também probabilística, no sentido de que ela reflete categorias – como os indivíduos da discutida “faixa intermediária” – de um contínuo gradual, as quais não estão nem em 0 nem em 1 somente, mas, sim, estão compreendidas nesse intervalo de 0 a 1, ora mais próximas de 0 (não-líder), ora mais próximas de 1 (líder).

Pela análise da TAB. 16, percebemos que, se inicialmente tínhamos 6 indivíduos líderes e 6 indivíduos não-líderes (Cf. capítulo 4), classificados assim segundo o critério de liderança de Labov (2001, v. 2), vemos que, em nossa pesquisa, pelo menos quando consideramos os três fenômenos em estudo ao mesmo tempo, esse quadro muda qualitativamente e numericamente. Há uma alteração qualitativa, no sentido de que um indivíduo que inicialmente fora classificado como líder, agora apresenta uma tendência de ser não-líder em nossa pesquisa. Este é o caso do indivíduo L3R com .49 (ver 3ª coluna da TAB. 16). Ainda, ocorreu de dois indivíduos não-líderes terem apresentado uma tendência de ser líder em nossa pesquisa. Este é o caso dos indivíduos NIL com .50 e NIA com .53 (ver 4ª coluna da TAB. 16). Essa alteração qualitativa gerou uma alteração quantitativa, visto que o número de não-líderes passou de 6 para 5 indivíduos (2ª e 3ª colunas da TAB. 16), enquanto que o número de líderes passou de 6 para 7 indivíduos (4ª e 5ª colunas da TAB. 16).

Pelo exposto nesta seção, nota-se que é importante considerar o indivíduo com unidade de análise (OLIVEIRA, M., 1992; SANGESTER, 2002). Dentro das categorias *líder* e *não-líder*, pode haver variabilidade entre os indivíduos (Cf. L3R, N1L e N1A na TAB. 16), o que faz com que estes, de certa forma, “destoem” de sua respectiva categoria.

Na próxima seção, iremos avaliar a relação entre as variáveis *Indivíduo* e *Fenômeno*, considerando os 3 fenômenos em estudo em conjunto e separadamente.

#### **5.2.5.1 Relação entre as variáveis *Indivíduo* e *Fenômeno***

Passamos agora a verificar a variável *Fenômeno* para cada um dos 12 indivíduos em análise, tendo em vista o fato de que esta tese se centra na investigação do comportamento do indivíduo (OLIVEIRA, M., 1992; DOCHERTY *et al.*, 1997; SANGSTER, 2002). Em tal análise, retomaremos a discussão do caráter de tendência de líderes e de não-líderes. Assim, por exemplo, investigaremos se, individualmente, quem tem tendência de líder é líder em todos os três fenômenos em estudo, ou se é não-líder em algum fenômeno específico. Para tanto, fazemos uma análise considerando todos os 3 fenômenos ao mesmo tempo e, posteriormente, cada um desses fenômenos em separado.

Na FIG. 18, é apresentada uma classificação em árvore (método CHAID) para os 12 indivíduos quanto aos três fenômenos em análise, ao mesmo tempo. Em tal árvore, vale lembrar que os líderes são indicados pela letra inicial “L” e os não-líderes, pela inicial “N” (Cf. QUADRO 6 da seção 4.3):

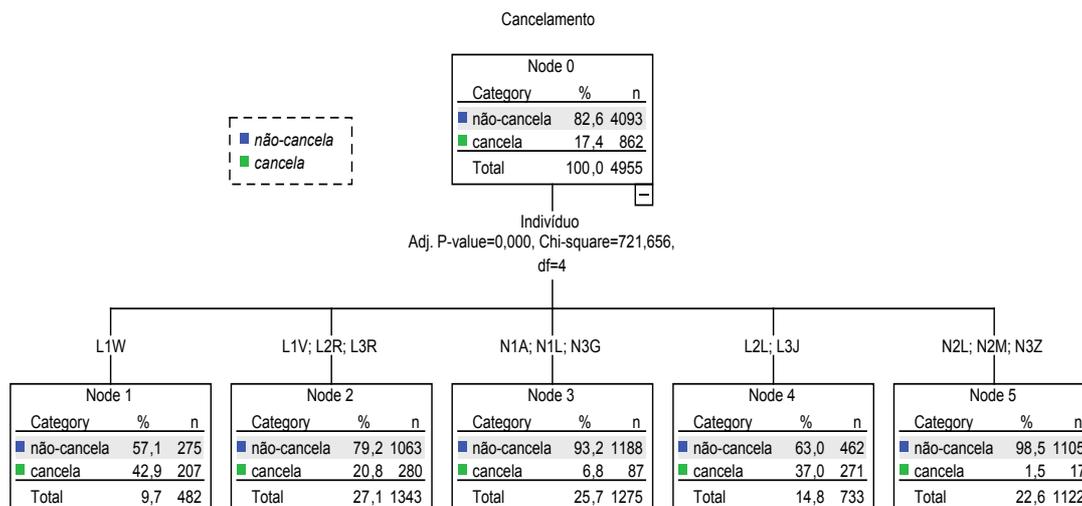


FIGURA 18 – Classificação em árvore (método CHAID) para os 12 indivíduos (6 líderes e 6 não-líderes) quanto aos três fenômenos em análise

Pela FIG. 18, percebemos que os 12 indivíduos deste estudo foram agrupados em 5 nós, em relação à variável dependente *Cancelamento*, considerando-se os três fenômenos aqui analisados. Como não há indivíduos líderes e indivíduos não-líderes juntos em um mesmo nó, mais uma vez (Cf. seção 5.2.3), podemos afirmar que, mesmo quando analisamos individualmente, em relação a esses três fenômenos em conjunto, líderes se comportam como líderes e não-líderes se comportam como não-líderes (LABOV, 2001, v. 2).

Contudo, por essa representação em árvore, podemos fazer uma observação. Os indivíduos líderes não se comportam de uma maneira única, já que, na FIG. 18, há três nós dividindo-os. O mesmo pode ser dito para os indivíduos não-líderes, pois em tal figura, há dois nós dividindo-os. Como não há apenas dois nós em tal árvore, um nó com os indivíduos líderes e o outro com os indivíduos não-líderes, essa divisão maior em cinco nós demonstra que existe diferença de comportamento tanto nos indivíduos

líderes entre si (1º, 2º e 4º nós da FIG. 18) quanto nos indivíduos não-líderes entre si (3º e 5º nós da FIG. 18). Assim, a FIG. 18 representa a variabilidade individual que há dentro das categorias *líder* e *não-líder*. Vale mencionar que isso não é contemplado na discussão que Labov (2001, v. 2) faz sobre os líderes da mudança.

Passemos, agora, à análise com todos os 12 indivíduos, contudo, para cada fenômeno em separado. Essa análise nos mostra que, dependendo do fenômeno, um indivíduo não-líder pode ser agrupado com indivíduos líderes e vice-versa. A FIG. 19, a qual considera somente o Fenômeno 1 (r #) (cancelamento de “r” final em nominais), já nos mostra isso:

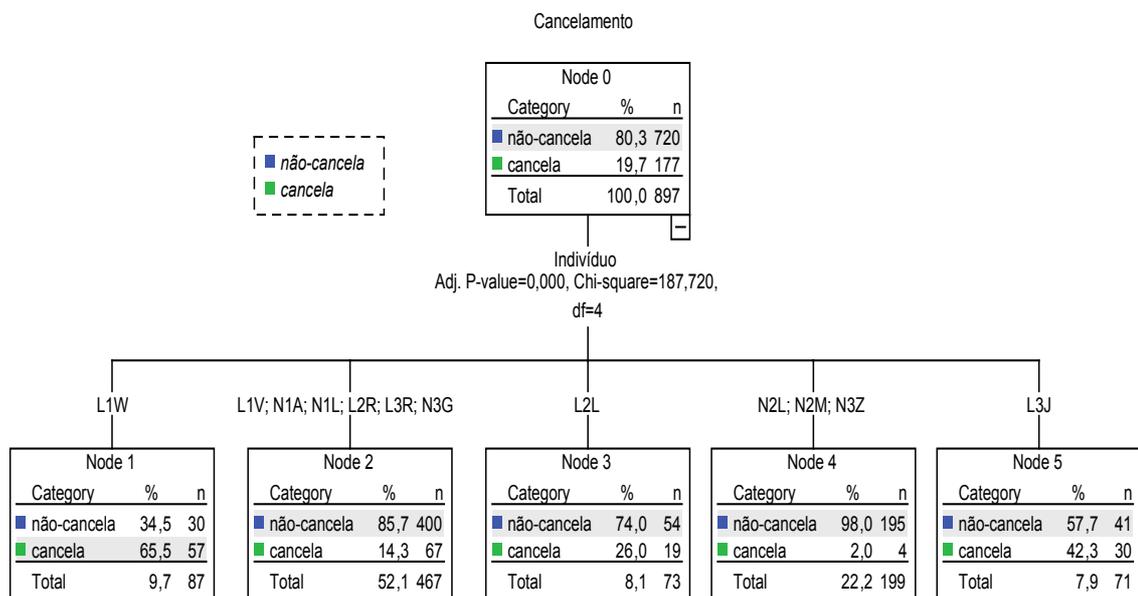


FIGURA 19 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 1 (r #) apenas

Como se pode verificar na árvore da FIG. 19, há indivíduos líderes e não-líderes agrupados em um mesmo nó (o 2º nó). Isso indica que, para o Fenômeno 1 (r #)

particularmente, há indivíduos líderes (L1V, L2R, L3R) e indivíduos não-líderes (N1A, N1L, N3G) que se comportam de forma parecida, ou seja, não há diferença estatisticamente significativa entre esses líderes e não-líderes com relação ao Fenômeno 1 (r #).

Esse fato aponta para a variabilidade individual e sugere que o perfil de líder (e, conseqüentemente, o de não-líder) é probabilístico (BOD; HAY; JANNEDY, 2003). Mesmo quando se considera um único indivíduo, este pode até apresentar uma tendência geral, por exemplo, de líder, contudo, a depender do fenômeno, pode também atingir um centro mais complexo dentro daquele contínuo de liderança representado na reta da TAB. 16, no qual ele ocupe um espaço que é típico de não-líder. Assim, no caso da reta representada na TAB. 16, esse indivíduo exemplificado, em geral líder, que tipicamente estaria mais à direita da reta – dentro da categoria líder (1) –, dependendo do fenômeno, estaria projetado mais à esquerda de tal reta – dentro da categoria não-líder (0). Vejamos agora o comportamento de todos os 12 indivíduos quanto ao Fenômeno 2 (v-v) (cancelamento de líquidas intervocálicas), na FIG. 20:

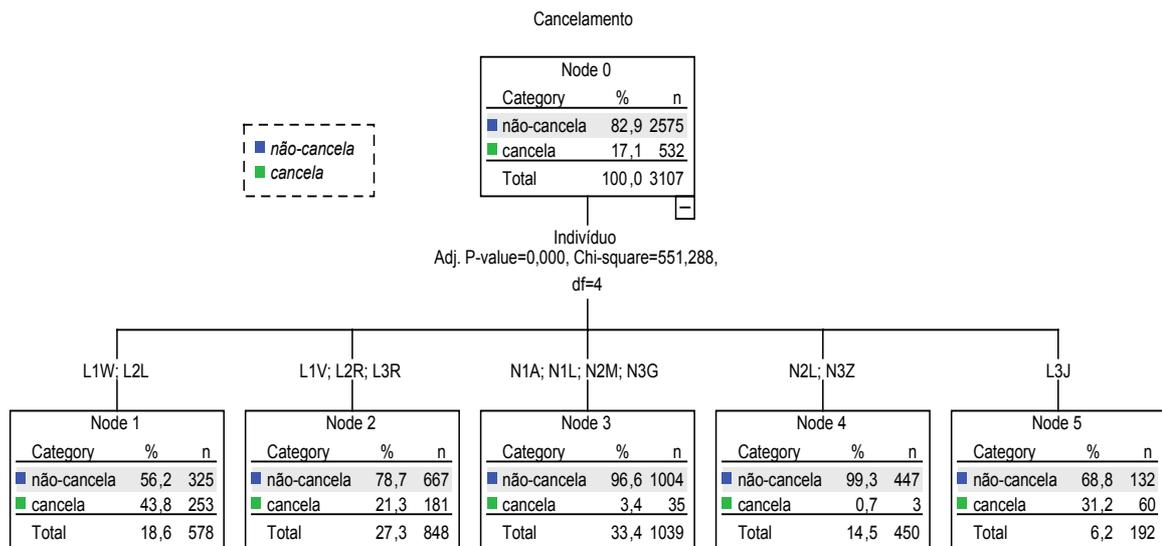


FIGURA 20 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 2 (v-v) apenas

Pela análise da árvore da FIG. 20, percebemos que, quanto ao Fenômeno 2 (v-v) somente, de modo diferente do Fenômeno 1 (r #) analisado anteriormente, não há indivíduos líderes e não-líderes agrupados em um mesmo nó. Isso indica que líderes e não-líderes se comportam de forma diferente quanto ao Fenômeno 2 (v-v), ou seja, a diferença entre eles é estatisticamente significativa. Líder se comporta como líder, e não-líder se comporta como não-líder quanto ao Fenômeno 2 (v-v) (LABOV, 2001, v. 2) Vejamos, por fim, o comportamento de todos os 12 indivíduos quanto ao Fenômeno 3 (cc) na FIG. 21:

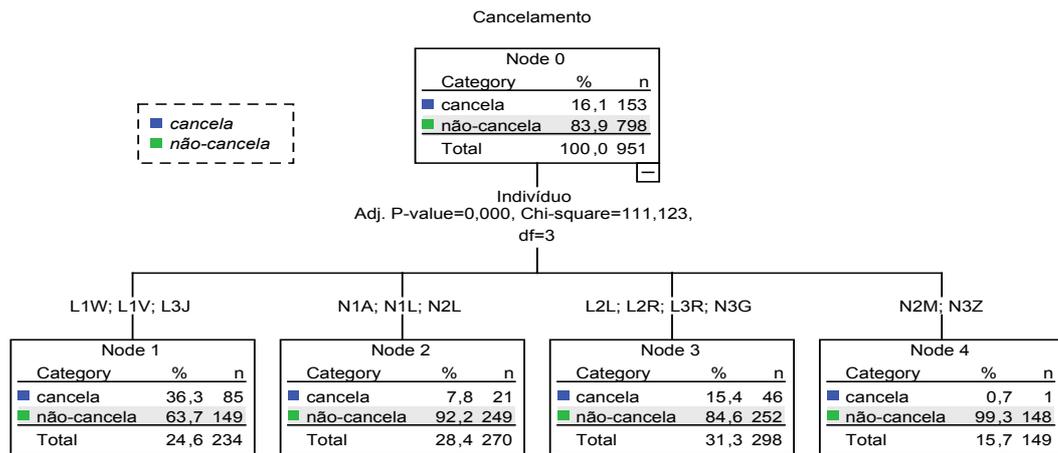


FIGURA 21 – Classificação em árvore (método CHAID) para todos os 12 indivíduos quanto ao Fenômeno 3 (cc) apenas

Pela análise da árvore da FIG. 21, a qual considera apenas o Fenômeno 3 (cc), percebemos que há um indivíduo não-líder, N3G, agrupado com outros três indivíduos líderes (L2L, L2R, L3R) no 3º nó. Isso indica que não houve diferença estatisticamente significativa entre tais indivíduos quanto ao Fenômeno 3 (cc), ou seja, que esse indivíduo não-líder (N3G) está, provavelmente, tendo um comportamento próximo de líder, por isso está em um nó onde predominam os líderes. Tal resultado também mostra que há variabilidade interindividual (SANGSTER, 2002) com relação à questão da liderança.

Novamente considerando a representação do contínuo de liderança da TAB. 16 (PIERREHUMBERT, 2002), quanto ao Fenômeno 3 (cc), o Indivíduo não-líder N3G estaria mais à direita da reta, na área de líder, portanto, mais próximo de 1 do que de 0. Contudo, quanto ao Fenômeno 2 (v-v) (Cf. FIG. 20), esse mesmo Indivíduo não-líder, N3G, estaria na área de não-líder mesmo, mais próximo de 0 do que de 1. Esse fato aponta para a existência de variabilidade intra-individual (SANGSTER, 2002), já que um mesmo indivíduo (N3G), de acordo com o fenômeno, ora apresentou

comportamento de não-líder (Fenômeno 2 (v-v)), ora apresentou comportamento de líder (Fenômeno 3 (cc)).

A fim de visualizar melhor os resultados desta seção, apresentamos o QUADRO 9 que sumariza as informações aqui discutidas:

**QUADRO 9: Líderes e não-líderes quanto aos fenômenos em análise**

Categoria	Todos os 3 fenômenos	Fenômeno 1		Fenômeno 2		Fenômeno 3	
		(r #)		(v-v)		(cc)	
Líder	L	L	N	L		L	N
Não-líder	N	L	N	N			N

No QUADRO 9, “L” significa líder e “N”, não-líder. Pela análise de tal quadro, vemos que, quanto a todos os fenômenos juntos (2ª coluna) e ao Fenômeno 2 (v-v) (4ª coluna), na categoria *líder* (1ª coluna), encontramos apenas líderes; enquanto que, na categoria *não-líder* (1ª coluna), encontramos somente não-líderes (LABOV, 2001, v. 2). Contudo, ao analisar, no quadro, os Fenômenos 1 (r #) e 3 (cc), vemos que não ocorre o mesmo. Quanto ao Fenômeno 1 (r #) (3ª coluna), tanto na categoria *líder* quanto na *não-líder*, há líderes e não-líderes juntos na mesma célula. Já quanto ao Fenômeno 3 (cc) (5ª coluna), se na categoria *não-líder*, há somente não-líderes, na categoria *líder*, há líderes e não-líderes. Esses fatos novamente apontam para a existência de variabilidade individual na análise de diferentes fenômenos, portanto, sinalizam a importância de se considerar o indivíduo como unidade de análise (OLIVEIRA, M., 1992).

Por todo o exposto nesta seção, concluímos que o comportamento do indivíduo deve ser avaliado em termos probabilísticos (BOD; HAY; JANNEDY, 2003), observando-se fenômenos específicos (SANGSTER, 2002). Realizamos uma análise de regressão para

os 12 indivíduos em estudo quanto aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc), separada e conjuntamente<sup>51</sup>, e obtivemos os seguintes resultados de peso relativo:

**TABELA 17: Resultados de peso relativo para os 12 indivíduos quanto aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em separado e em conjunto**

Indivíduos	Peso relativo Fenômeno 1 (r #)	Peso relativo Fenômeno 2 (v-v)	Peso relativo Fenômeno 3 (cc)	Peso relativo todos os fenômenos
Indivíduo 1 – L1W	.93	.88	.93	.94
Indivíduo 2 – L1V	.57	.71	.96	.57
Indivíduo 3 – N1A	.55	.25	.74	.53
Indivíduo 4 – N1L	.52	.29	.78	.50
Indivíduo 5 – L2L	.71	.91	.85	.70
Indivíduo 6 – L2R	.58	.79	.84	.60
Indivíduo 7 – N2L	.15	.08	.59	.15
Indivíduo 8 – N2M	.10	.21	.00	.11
Indivíduo 9 – L3R	.55	.75	.86	.49
Indivíduo 10 – L3J	.84	.83	.95	.85
Indivíduo 11 – N3Z	.11	.06	.30	.13
Indivíduo 12 – N3G	.45	.34	.87	.45

Na 1ª coluna da TAB. 17, os indivíduos líderes estão destacados de cinza escuro, enquanto que os não-líderes não apresentam qualquer destaque. Nas 2ª, 3ª e 4ª colunas de tal tabela, as células destacadas em cinza claro mostram que fenômeno cada indivíduo apresenta maior probabilidade de realizar. Assim, vemos que, com exceção total do Indivíduo 8 – N2M, que apresenta 00., todos os outros indivíduos apresentam maior probabilidade de realizar o Fenômeno 3 (cc), sendo que o Indivíduo 1 – L1W favorece mais a realização dos Fenômenos 1 (r #) e 3 (cc), de forma igual, com .93.

O fato de o Fenômeno 3 (cc) ser o que apresenta maior probabilidade de realização para a grande maioria dos indivíduos pode estar relacionado, individualmente, à palavra *outro* e não a todas as palavras de tal fenômeno aqui avaliadas. Inclusive, como

<sup>51</sup> Os resultados de peso relativo para todos os três fenômenos em conjunto, presentes na última coluna da TAB. 17, foram discutidos por ocasião da análise da TAB. 9, na seção 5.2.2.

analisamos na seção 5.2.2, que tratou da variável *Palavra*, dentre todas as 30 palavras dos 3 fenômenos em análise, *outro* foi a que apresentou maior percentual e maior probabilidade de cancelamento, com, respectivamente, 39,8% e .91 (Cf. TAB. 9). Ainda, *outro* se mostrou estatisticamente diferente de todas as demais palavras em estudo, por ter ficado sozinha em um nó, na árvore da FIG. 6 (Cf. seção 5.2.2). Em função disso, notamos que aqui, mais uma vez, é essencial considerar o modelo da Difusão Lexical (CHEN; WANG, 1975) e, conseqüentemente, a Fonologia de Uso (BYBEE, 2001), os quais dão destaque à palavra na implementação da mudança sonora, postulando que esta ocorre de modo lexicalmente gradual.

Uma outra observação que podemos fazer a respeito da TAB. 17, é que há indivíduos que apresentam valores próximos de peso relativo para os três fenômenos, somente favorecendo (Indivíduos 1, 2, 5, 6, 9 e 10 – todos líderes) ou somente desfavorecendo (Indivíduos 8 e 11, ambos não-líderes). Por exemplo, o Indivíduo 1 – L1W, que apresenta os valores, .93, .88 e .93 para os Fenômenos 1 (r #), 2 (v-v) e 3 (cc), respectivamente, só favorece. Já o Indivíduo 8 – N2M, com .10, .21, .00, só desfavorece. Com isso, podemos dizer que, enquanto L1W se comporta como líder em relação a todos os três fenômenos, N3M se comporta como não-líder em relação a eles (SANGSTER, 2002; LABOV, 2001, v. 2). Dito de outra forma, retomando a representação da TAB. 16, na gradação que vai da categoria não-líder (0) à categoria líder (1) (PIERREHUMBERT, 2001), enquanto L1W estaria bem próximo de 1, N3M estaria bem próximo de 0, com relação aos três fenômenos em estudo.

Por outro lado, os outros indivíduos (3, 4, 7 e 12 – todos não-líderes) apresentam valores distantes de peso relativo para os três fenômenos. Por exemplo, o Indivíduo 3 – N1A que favorece o cancelamento quanto a dois fenômenos (Fenômeno 1 (r #): .55;

Fenômeno 3 (cc): .74), desfavorece quanto a um fenômeno (Fenômeno 2 (v-v): .25). Esse resultado aponta para uma questão discutida ainda nesta seção: quanto à liderança (LABOV, 2001, v. 2), alguns indivíduos se comportam como líder ou não-líder de acordo com o fenômeno. Assim, vemos que o perfil de líder (ou de não-líder) pode não ser constante e invariável. Isso é o que vemos com relação ao Indivíduo 3 – N1A, que apresenta comportamento de líder quanto aos Fenômenos 1 (r #) e 3 (cc), por favorecê-los, mas apresenta comportamento de não-líder quanto ao Fenômeno 2 (v-v), por desfavorecê-lo.

Tal achado também aponta para o fato de que é importante analisar fenômenos diferentes em separado para um mesmo indivíduo, pois isso pode permitir a observação de variações intra-individuais (SANGSTER, 2002; DOCHERTY *et al.*, 1997), como as que constatamos com relação aos últimos indivíduos citados, 3, 4, 7 e 12. Como se apontou no capítulo 1, geralmente, os trabalhos em lingüística avaliam um único fenômeno para todos os indivíduos avaliados. Este é o caso dos trabalhos referentes aos corpora aqui considerados: Fontes Martins (2001), Cristófaró Silva (2003) e Huback (2003). Como esta pesquisa analisa três fenômenos, ao mesmo tempo e em separado, para cada indivíduo em estudo, pudemos perceber a importância de se avaliar fenômenos diferentes para tais indivíduos (SANGSTER, 2002), a fim de se observar variações inter e intra-individuais. Esta seria até uma proposta metodológica desta pesquisa para trabalhos futuros que se ocupem em avaliar o comportamento do indivíduo quanto ao componente fonológico.

Ainda com relação à TAB. 17, é interessante observar que, mesmo indivíduos que apresentam comportamento só de líder (só favorecendo os fenômenos de cancelamento) ou só de não-líder (desfavorecendo tais fenômenos) apresentam valores de peso relativo,

ou seja, probabilidades que se distanciam. Assim, o Indivíduo 2 – L1V, por exemplo, favorece os três fenômenos, mas com probabilidades distantes: Fenômeno 1 (r #): .57; Fenômeno 2 (v-v): .71; Fenômeno 3 (cc): .96. Ainda, o Indivíduo 8 – N2M, que desfavorece os três fenômenos, apresenta probabilidades, de certa forma, distantes: Fenômeno 1 (r #): .10; Fenômeno: .21; Fenômeno: .00.

Esse fato nos mostra que indivíduos lidam com diferentes fenômenos de forma probabilística (BOD; HAY; JANNEDY, 2003). Mesmo que um indivíduo, por exemplo, sempre favoreça a aplicação de fenômenos de mudança sonora – como é o caso do Indivíduo 2 – L1V citado – ele o fará de forma gradiente (PIERREHUMBERT, 2002). Podemos novamente retomar a representação da TAB. 16, para criar a FIG. 22 e observar que, com relação ao contínuo de 0 – não-líder a 1 – líder, o Indivíduo 2 estaria projetado à direita da reta, em direção a 1 – líder, contudo em pontos diferentes de acordo com o fenômeno (ver FIG. 22 abaixo). Este achado confirma o caráter probabilístico da linguagem (BOD; HAY; JANNEDY, 2003).

### Indivíduo 2 – L1V

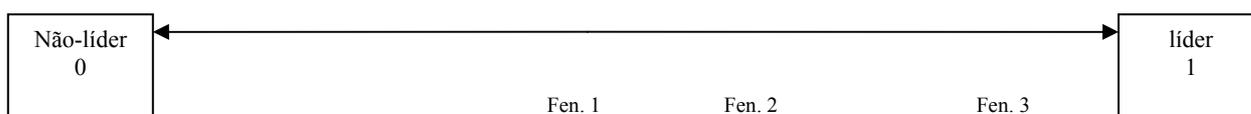


FIGURA 22 – Gradiência de fenômenos com relação ao Indivíduo 2 – L1V

De modo geral, os resultados desta seção, referentes à relação entre as variáveis *Indivíduo* e *Fenômeno*, evidenciam que a organização do componente fonológico pode variar de indivíduo para indivíduo a depender do fenômeno. Na próxima seção, avaliamos a relação da variável *Indivíduo* com a variável *Palavra*, a fim de verificar

como cada um dos 12 indivíduos em análise se relacionam com as 30 palavras aqui avaliadas.

#### **5.2.5.2 Relação entre as variáveis *Indivíduo* e *Palavra***

A fim de avaliarmos a relação da variável *Indivíduo* com a variável *Palavra*, apresentamos, da FIG. 23 a FIG. 34, análises de classificação em árvore (método CHAID) para os 12 indivíduos deste trabalho quanto às 30 palavras do estudo transversal<sup>52</sup>. Tais análises demonstram, para cada indivíduo, como as palavras se organizam em grupos/ nós, de acordo com a homogeneidade que há entre elas, em termos da média percentual de cancelamento.

---

<sup>52</sup> Gostaríamos de ter realizado análises de regressão logística para abordar a relação das variáveis *Indivíduo* e *Palavra*. Contudo, como os dados se encontravam mal distribuídos, o programa SPSS não aceitou rodar tais análises. Ainda assim, a fim de demonstrar com clareza os resultados desta seção, no APÊNDICE F, apresentamos tabelas com os percentuais de cancelamento, no estudo transversal, de cada indivíduo para cada uma das 30 palavras. Vale destacar que, pelas análises do APÊNDICE F, vemos que um mesmo indivíduo pode ter produções diferentes para uma única palavra. Por exemplo, o Indivíduo – L1W apresenta 11 ocorrências para a palavra *melhor*, sendo que enquanto em 6 delas, ele pronuncia o “r” final, em 5, ele não o pronuncia, cancelando-o. Esse fato demonstra variabilidade intra-individual, o que se contrapõe ao estudo de Oliveira, M. (2006).

## Líderes do Corpus 1:

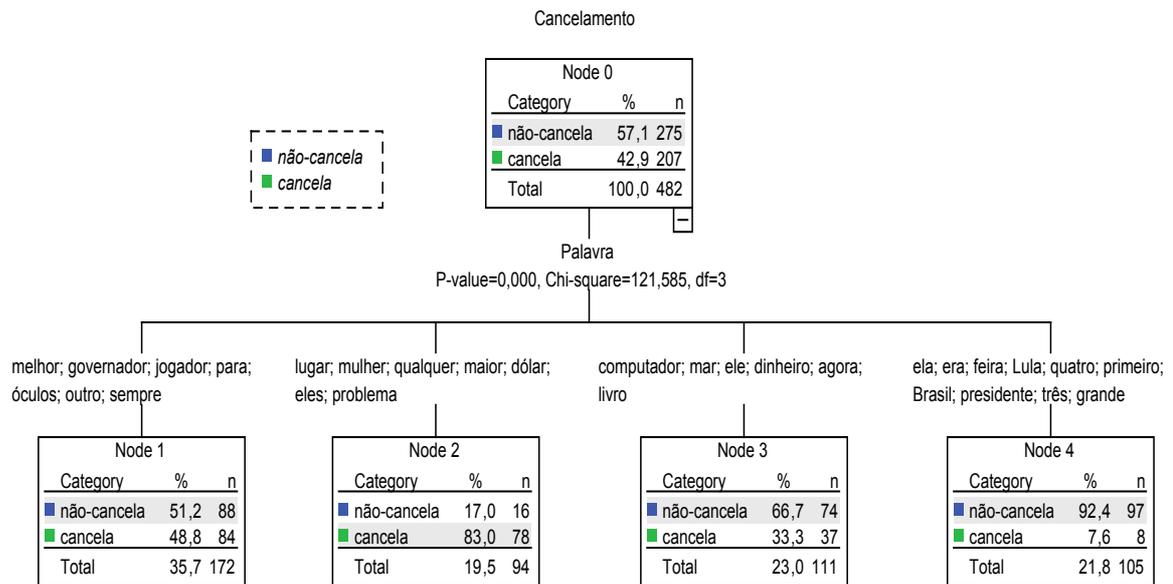


FIGURA 23 – Indivíduo 1 – L1W e as 30 palavras em estudo

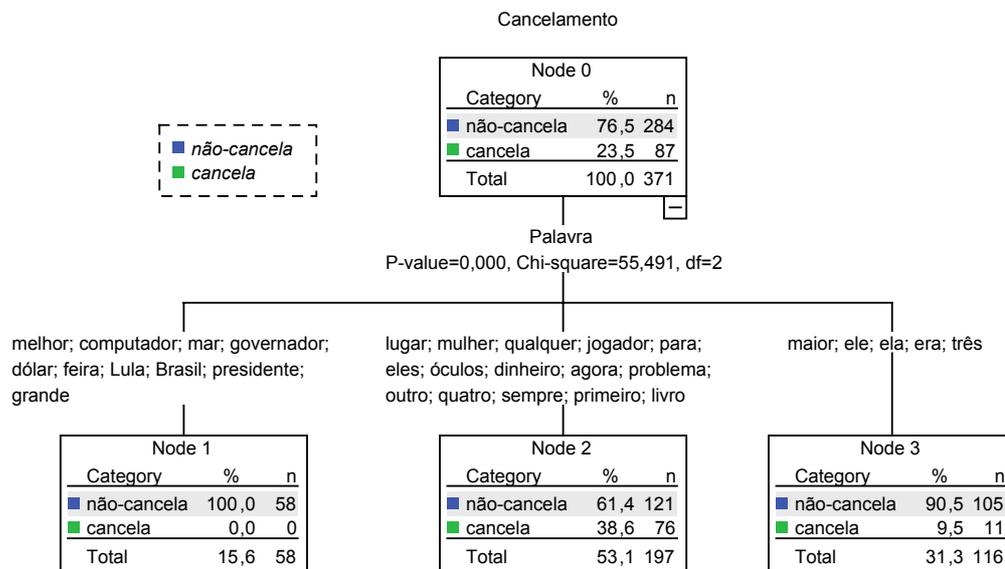


FIGURA 24 – Indivíduo 2 – L1V e as 30 palavras em estudo

### Não-líderes do Corpus 1:

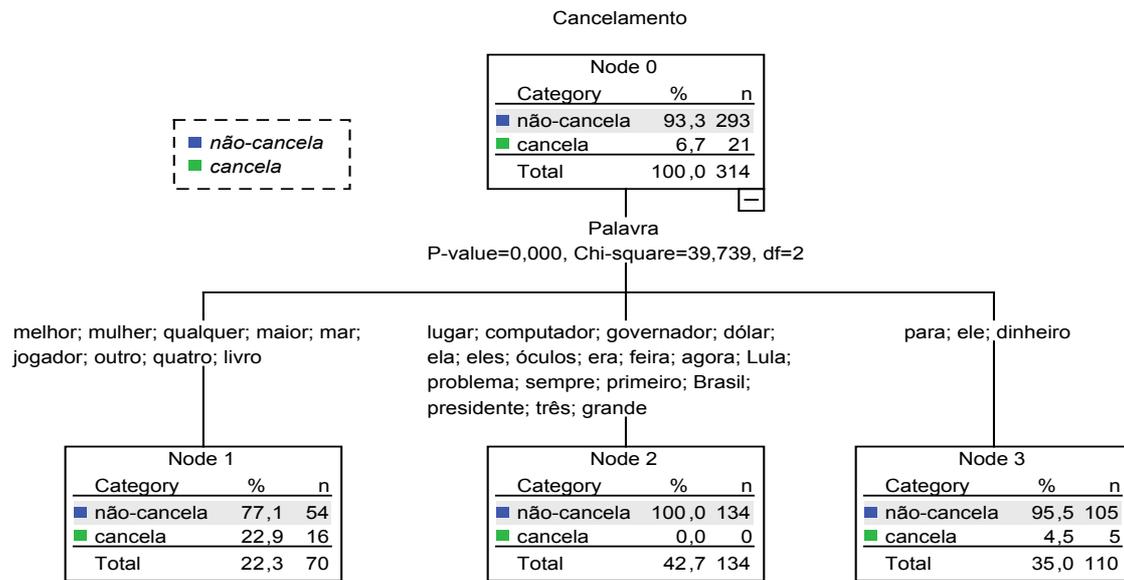


FIGURA 25 – Indivíduo 3 – N1A e as 30 palavras em estudo

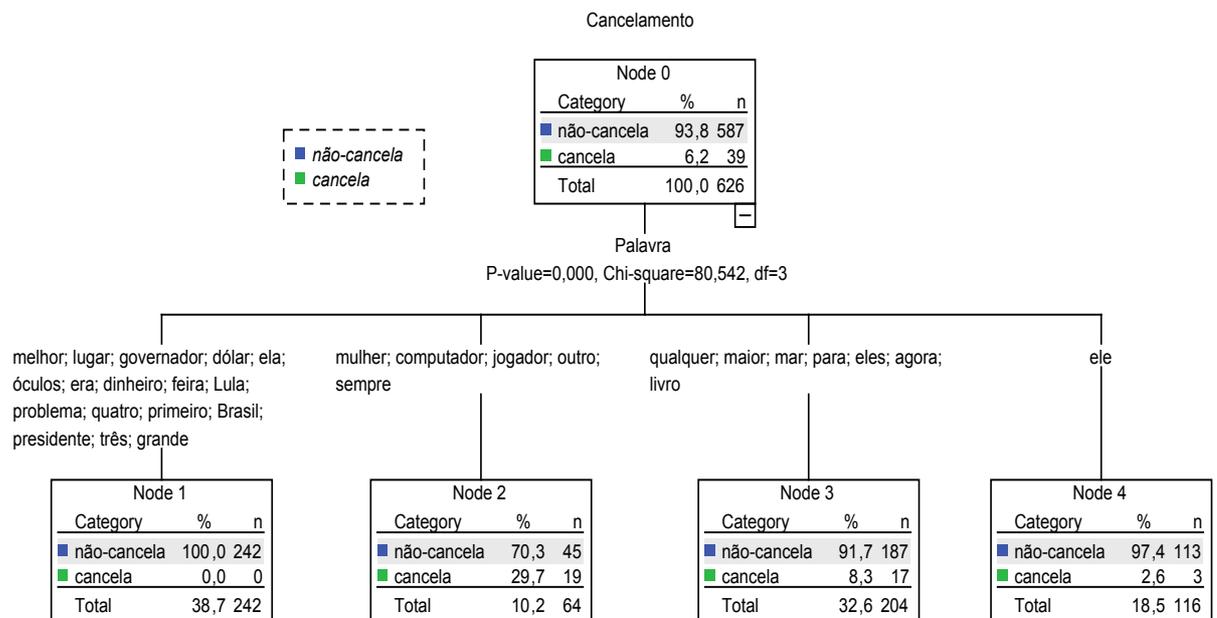


FIGURA 26 – Indivíduo 4 – N1L e as 30 palavras em estudo

## Líderes do Corpus 2:

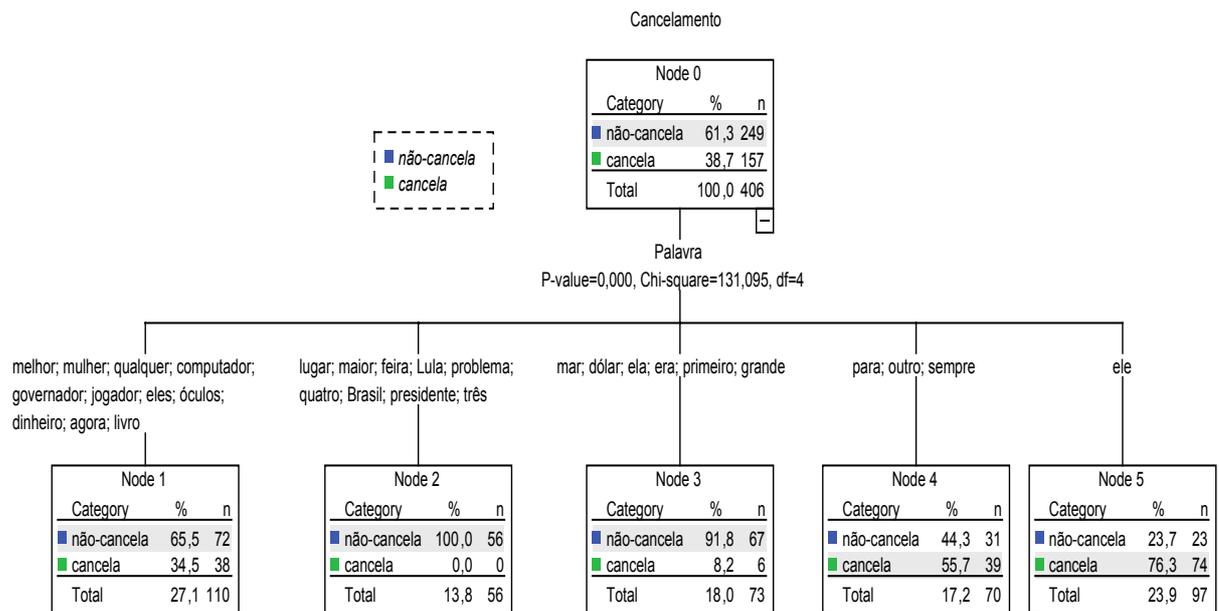


FIGURA 27 –Indivíduo 5 – L2L e as 30 palavras em análise

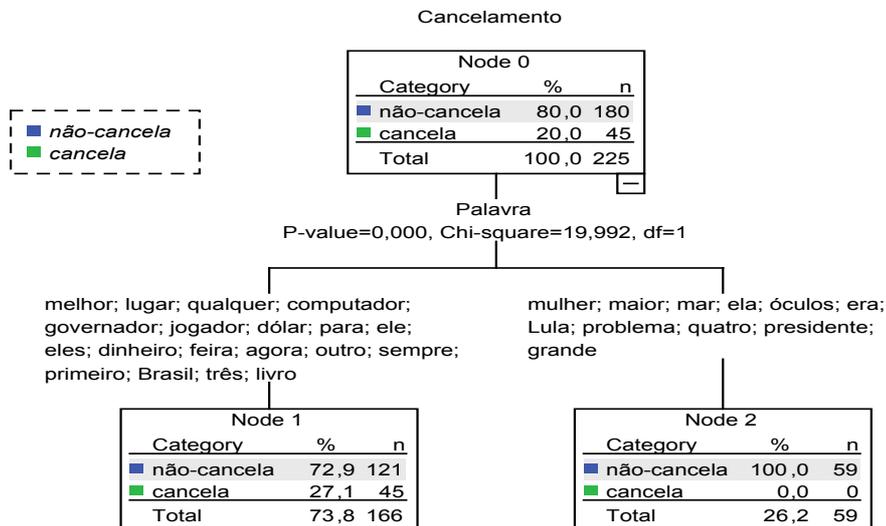


FIGURA 28 – Indivíduo 6 – L2R e as 30 palavras em estudo

**Não-líderes do Corpus 2:**

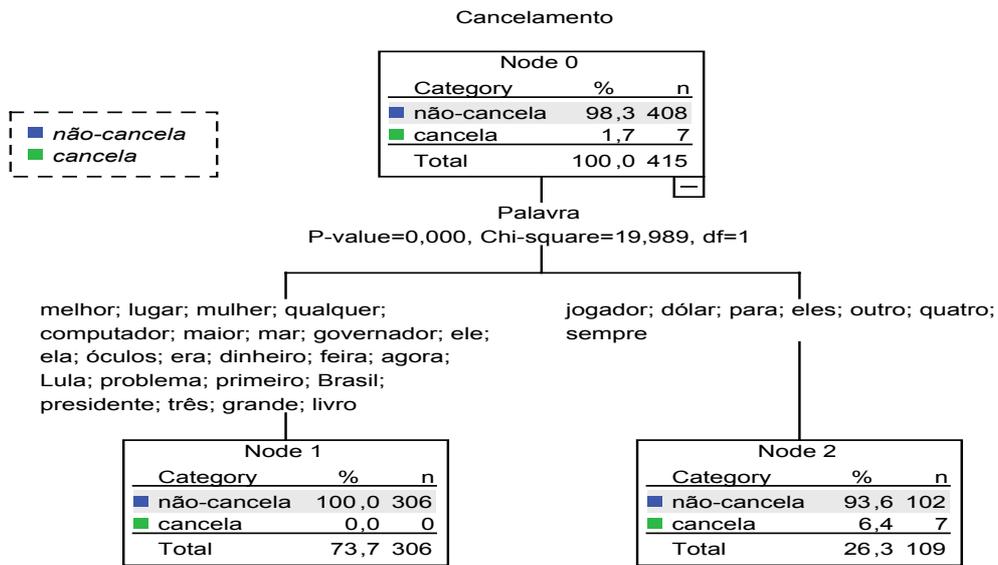


FIGURA 29 – Indivíduo 7 – N2L e as 30 palavras em estudo

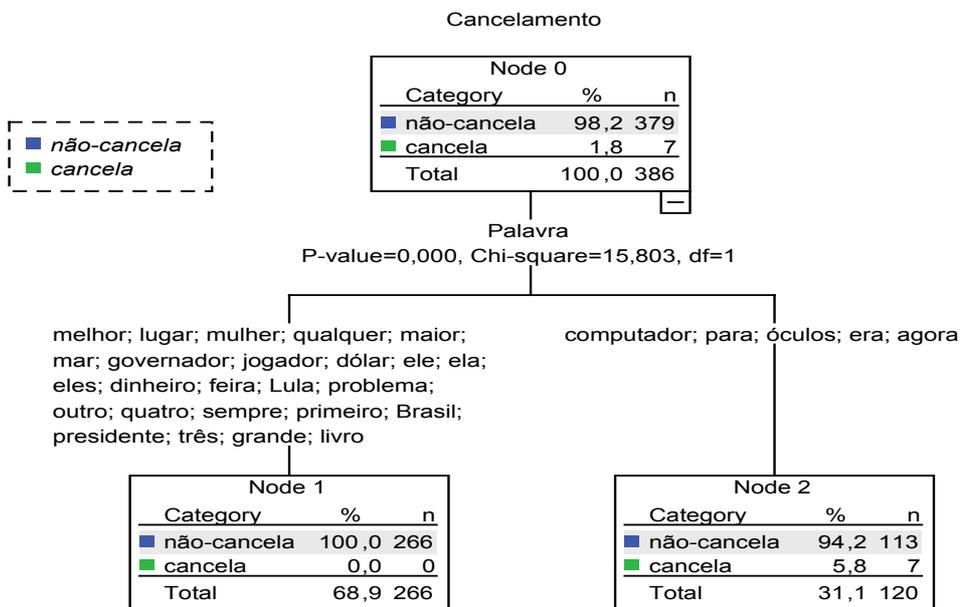


FIGURA 30 – Indivíduo 8 – N2M e as 30 palavras em estudo

Líderes do Corpus 3:

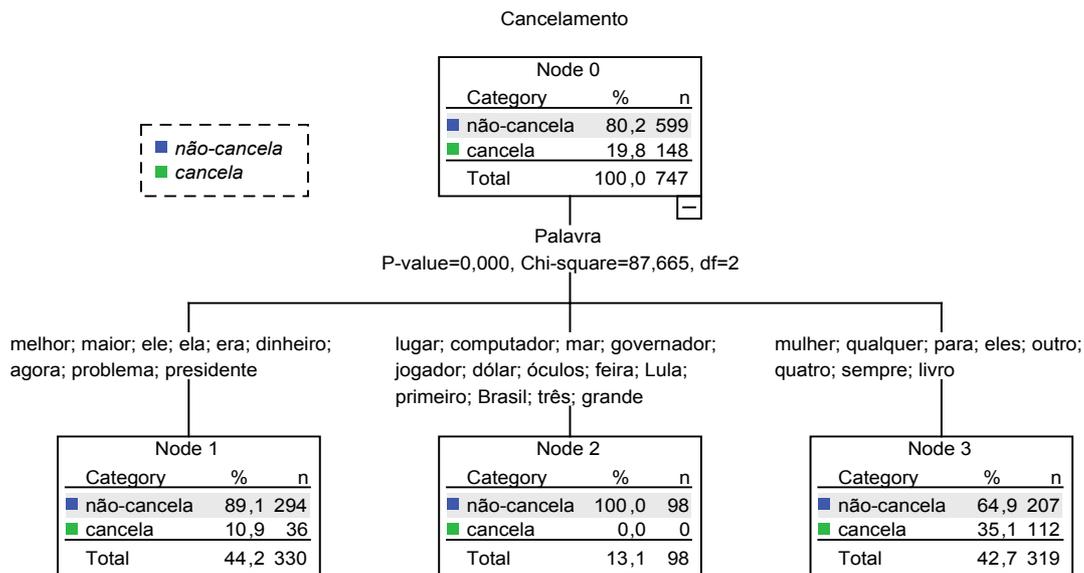


FIGURA 31 – Indivíduo 9 – L3R e as 30 palavras em estudo

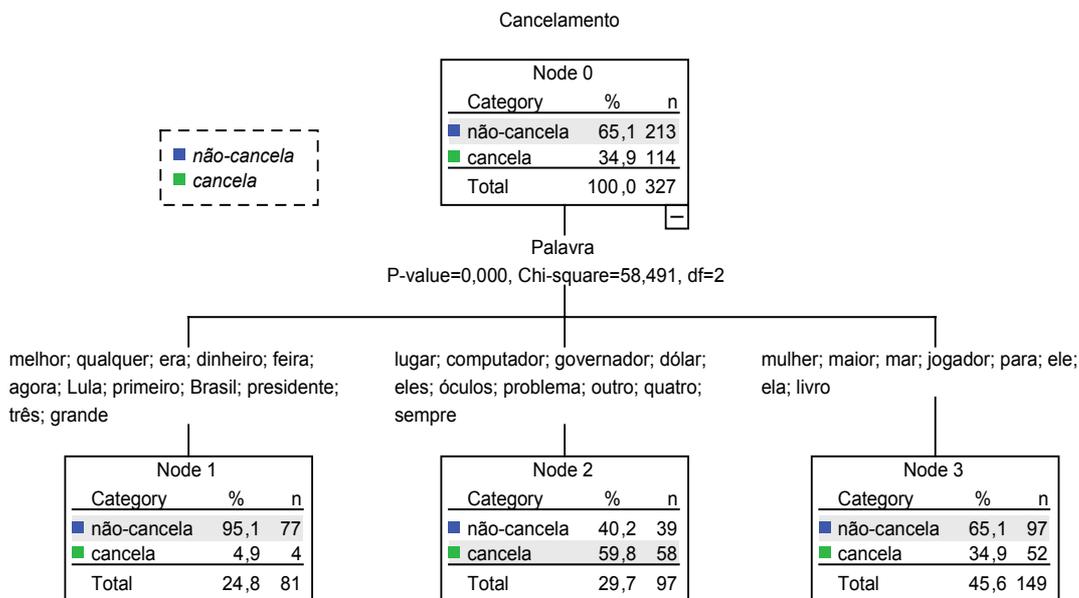


FIGURA 32 – Indivíduo 10 – L3J e as 30 palavras em estudo

### Não-líderes do Corpus 3:

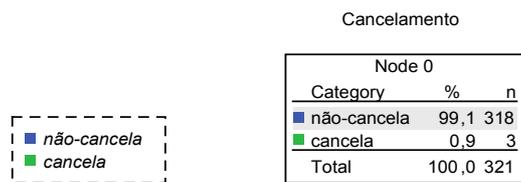


FIGURA 33 – Indivíduo 11 – N3Z e as 30 palavras em estudo

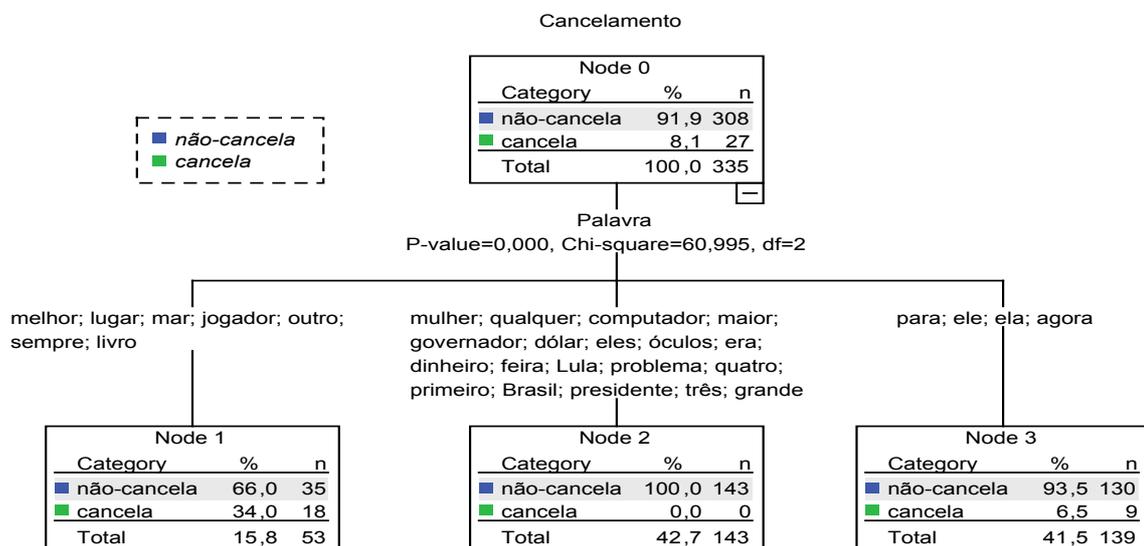


FIGURA 34 – Indivíduo 12 – N3G e as 30 palavras em análise

Analisando as árvores das FIG. 23 a 34, ressalta a ocorrência de grande variabilidade interindividual (OLIVEIRA, M., 1992; SANGSTER, 2002). Isso porque, mesmo quando avaliamos somente líderes ou somente não-líderes provenientes de um mesmo corpus, vemos que há entre os indivíduos, grandes diferenças na forma de agrupar as 30 palavras deste estudo.

Assim, tomando como exemplo os líderes do Corpus 1, enquanto o Indivíduo 1 – L1W agrupa as 30 palavras em 4 nós (FIG. 23), o Indivíduo 2 – L1V as agrupa em 3 nós (FIG. 24). Ainda, os percentuais de cancelamento médio dos nós para cada um desses indivíduos variam muito: enquanto os quatro nós de L1W variam de 7,6% a 83% de cancelamento (4º e 2º nós, respectivamente, da FIG. 23), os três nós de L1V variam de 0% a 38,8% de cancelamento (1º e 2º nós, respectivamente, da FIG. 24). Como um exemplo de não-líderes, vale citar o caso dos indivíduos do Corpus 3. Enquanto o Indivíduo 12 – N3G agrupou as 30 palavras em 3 nós, que variaram de 0% a 34% de cancelamento (FIG. 34), o Indivíduo 11 – N3Z apresentou um único nó com 0,9% de cancelamento (FIG. 33). N3Z foi o único indivíduo de nosso corpus que não apresentou diferença estatisticamente significativa entre as palavras em análise, de modo que elas ficaram agrupadas em um mesmo nó.

Em todos os outros indivíduos deste estudo (Cf. FIG. 25 a 32), observamos variações como as exemplificadas acima. É importante destacar que a variabilidade interindividual na forma de lidar com as palavras, encontrada em nossos dados, é indício de que o modo de representar e organizar o léxico varia entre os indivíduos e ocorre de forma probabilística (BYBEE, 2001; BOD; HAY; JANNEDY, 2003; REINECKE, 2006).

Por fim, corroborando os resultados da seção 5.2.2 especialmente, vemos que, nas análises de todos os indivíduos (Cf. FIG. 23 a FIG. 34), as palavras não aparecem agrupadas nos nós, em função dos seus respectivos fenômenos. Assim, por exemplo, na citada análise do Indivíduo 2 – L1V (FIG. 24), no 1º nó de sua árvore em que estão agrupadas as palavras que tiveram o menor percentual médio de cancelamento (0%), vemos que há palavras relacionadas aos três fenômenos aqui avaliados: *melhor*,

*computador, mar, governador, dólar* (Fenômeno 1 (r #)); *feira, Lula* (Fenômeno 2 (v-v)); *Brasil, presidente, grande* (Fenômeno 3 (cc)). Ainda considerando esse mesmo indivíduo (FIG. 24), no 2º nó de sua árvore em que estão agrupadas as palavras que tiveram o maior percentual médio de cancelamento (38,6%), também vemos palavras relacionadas aos três fenômenos: *lugar, mulher, qualquer, jogador* (Fenômeno 1 (r #)); *para, eles, óculos, dinheiro, agora* (Fenômeno 2 (v-v)); *problema, outro, quatro, sempre, primeiro, livro* (Fenômeno 3 (cc)). Esse fato é evidência para a idéia de que a mudança sonora se difunde de modo gradual no léxico (CHEN; WANG, 1975; BYBEE, 2001).

Também os resultados desta seção, de modo geral, nos mostram que a organização do componente fonológico varia individualmente em função da variável *Palavra*. Indivíduos diferentes apresentam léxicos diferentes (o que constatamos pelas representações em árvore da FIG. 23-34) e, como consequência, emergências de gramática diferentes, tendo em vista que a gramática emerge de conexões lexicais (HOPPER, 1987; BYBEE, 2001). Passemos agora à última análise que fazemos para o estudo transversal, relacionando as variáveis *Indivíduo* e *Estilo*, a fim de verificarmos possíveis variações inter e intra-individuais decorrentes do fator estilo (SANGSTER, 2002).

### **5.2.5.3 Relação entre as variáveis *Indivíduo* e *Estilo***

Para investigar como cada um dos 12 indivíduos deste estudo se comporta nos 3 estilos aqui avaliados, propomos as análises de classificação em árvore (método CHAID) das

FIG. 35, FIG. 36 e FIG. 37. Estas se relacionam, respectivamente, aos estilos *fala espontânea*, *tarefa de nomeação por figura* e *tarefa de leitura*. Começemos pela FIG. 35:

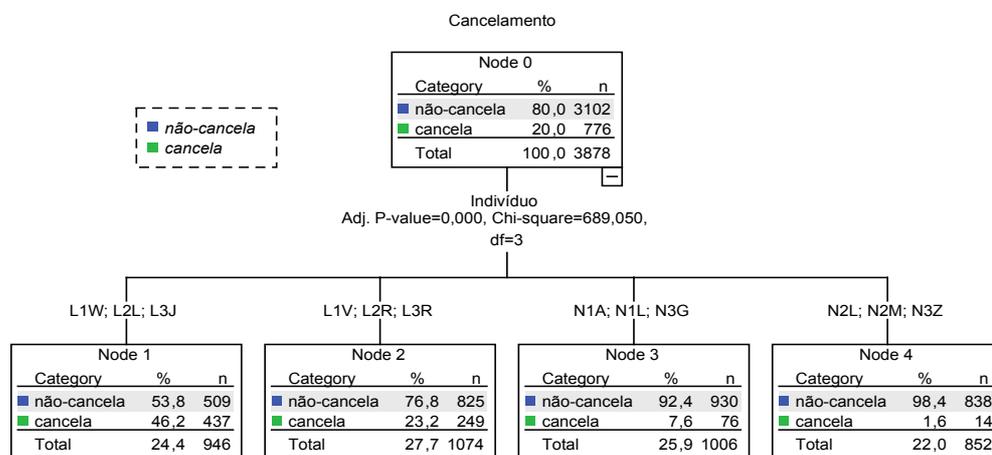


FIGURA 35 – Os 12 indivíduos no estilo fala espontânea

A FIG. 35 nos mostra que, quanto ao estilo *fala espontânea*, os 12 indivíduos foram divididos em 4 nós, sendo que: os 2 primeiros são compostos por líderes (um nó com uma média de 46,2% de cancelamento e ou outro com 23,2%); e os 2 últimos compostos são por não-líderes (um nó com 7,6% de cancelamento e o outro com 1,6%).

O que vemos na FIG. 35 é que, quanto aos líderes, os indivíduos do 1º nó (L1W, L2L e L3J) apresentam praticamente o dobro de percentual médio de cancelamento dos indivíduos do 2º nó (L1V, L2R e L3R). Já quanto aos não-líderes, os indivíduos do 3º nó (N1A, N1L, N3G) apresentam uma média percentual de cancelamento, aproximadamente, 4 vezes maior do que os indivíduos do 4º nó (N2L, N2M, N3Z). Esses achados apontam para a existência de variabilidade interindividual na avaliação de um único estilo (SANGESTER, 2002; DOCHERTY *et al.*, 1997), tendo em vista que os percentuais médios de cancelamento dos líderes entre si e dos não-líderes entre si

apresentam diferença estatisticamente relevante na *fala espontânea*. Passemos agora ao estilo *tarefa de nomeação por figura* com a FIG. 36:

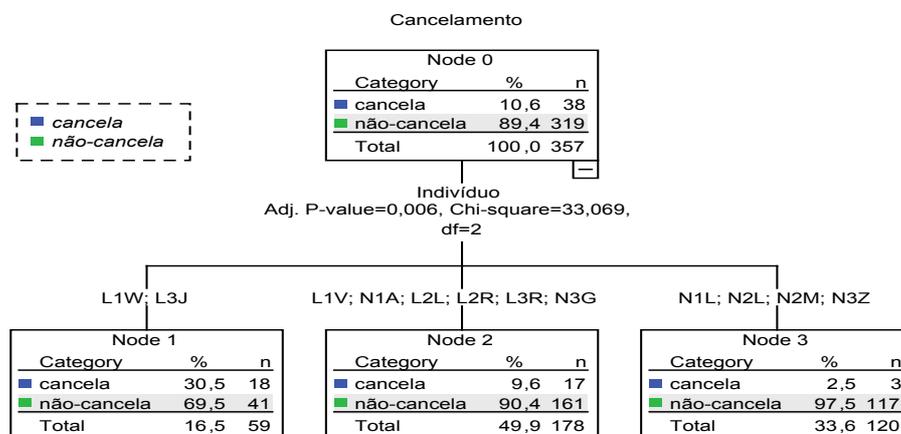


FIGURA 36 – Os 12 indivíduos no estilo tarefa de nomeação de figura

Na árvore da FIG. 36 referente ao estilo *nomeação de figura*, os indivíduos foram divididos em um nó a menos do que na árvore da FIG. 35 referente à *fala espontânea*, ou seja, foram divididos em 3 nós. No 1º nó da árvore da FIG. 36, que apresenta o maior percentual médio de cancelamento (30,5%), há dois indivíduos líderes, L1W e L3J. No 2º nó de tal árvore, o qual tem um percentual médio de cancelamento intermediário (9,6%), ocorre algo diferente: há indivíduos líderes (L1V, L2L, L2R e L3R) e não-líderes (N1A e N3G). Por fim, o 3º nó, que apresenta o menor percentual médio de cancelamento (2,5%), há somente não-líderes (N1L, N2L, N2M e N3Z). O fato de haver o 2º nó com uma mistura de líderes e não-líderes, novamente, aponta para a variabilidade interindividual em um único estilo (SANGSTER, 2002), por demonstrar que há diferença estatisticamente relevante nos líderes entre si e nos não-líderes entre si, na *nomeação por figura*.

Deve-se destacar que o Indivíduo – L2L que se encontra no grupo dos indivíduos com maior percentual médio de cancelamento (46,2%, Cf. 1º nó da FIG. 35) na *fala*

*espontânea*, encontra-se no grupo dos indivíduos líderes e não-líderes com o percentual médio de cancelamento intermediário (9,6%, Cf. 2º nó da FIG. 36) na *nomeação por figura*. Ainda, o Indivíduo – N1L que se encontra no grupo dos não-líderes com maior percentual de cancelamento (7,6%, Cf. 3º nó da FIG. 35) na *fala espontânea*, encontra-se no grupo de líderes e não-líderes com um percentual médio de cancelamento intermediário (9,6%, Cf. 2º nó da FIG. 36) na *nomeação por figura*. Esse resultado demonstra variabilidade intra-individual decorrente do fator estilo (SANGSTER, 2002), já que um mesmo indivíduo (no caso, os citados L2L e N1L) apresenta comportamentos lingüísticos diferentes de acordo com o estilo. Vejamos agora a FIG. 37 que se refere ao estilo *tarefa de leitura*:

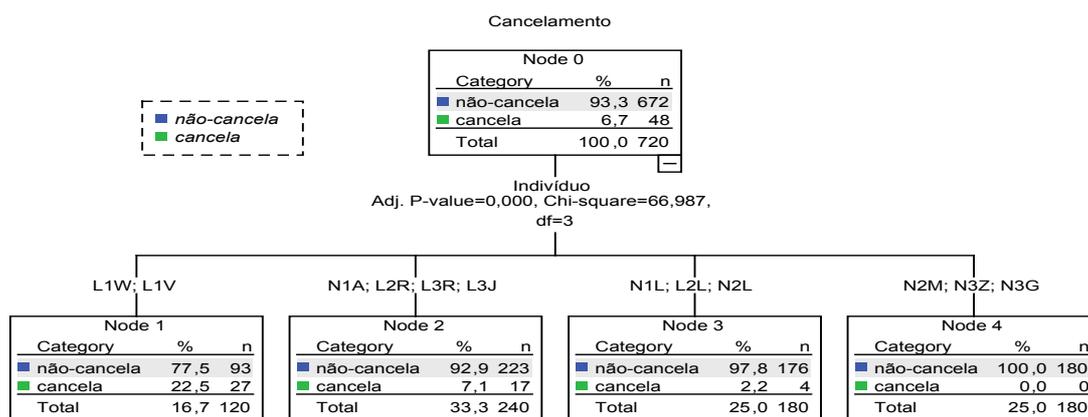


FIGURA 37 – Os 12 indivíduos no estilo tarefa de leitura

Pela análise da FIG. 37 que se refere à *leitura*, notamos que, assim como na análise da *fala espontânea* (FIG. 35), os 12 indivíduos foram divididos em 4 nós. Contudo, há uma diferença entre as análises desses dois estilos: à semelhança do que ocorreu na *nomeação por figura* (FIG. 36), há dois nós, na *leitura* (FIG. 37), que apresentam líderes e não-líderes ao mesmo tempo – 2º nó (N1A, L2R, L3R e L3J) e 3º nó (N1L, L2L e N2L). Estes dois nós – o 2º com 7,1% de cancelamento e o 3º com 2,2% –

apresentam um percentual médio de cancelamento intermediário em relação aos outros nós, o 1º e o 4º, que apresentam, respectivamente, 22,5% e 0% de cancelamento. Novamente, a existência desses grupos em que há líderes e não-líderes simultaneamente demonstra que pode haver variabilidade interindividual em um mesmo estilo, no caso a *leitura* (SANGSTER, 2002). Também aqui notamos variações intra-individuais ao comparar as árvores das FIG. 35, FIG. 36 e FIG. 37: a) os indivíduos líderes L1V, L3J, L2L, a depender do estilo, encontram-se ou no grupo de maior percentual médio de cancelamento, ou no grupo de um percentual médio de cancelamento intermediário; b) os indivíduos não-líderes N1L, N2L e N3G, também dependendo do estilo, encontram-se ou no grupo com menor percentual médio de cancelamento, ou no grupo de um percentual médio de cancelamento intermediário.

Na análise desta seção, destacamos que os resultados nos levam a formular, mais uma vez, que a organização do componente fonológico pode variar, individualmente, de acordo com o estilo (SANGSTER, 2002). Terminamos, assim, a análise do estudo transversal aqui realizado. Contudo, a avaliação do comportamento do indivíduo continua no próximo capítulo em que abordamos uma análise de caráter longitudinal e o estudo longitudinal aqui realizado propriamente.

### **5.3 Sumário**

Neste capítulo, abordamos o estudo transversal aqui feito com os 12 participantes, no decorrer da seção 5.2, em que apresentamos e discutimos os resultados referentes aos dados de tal estudo. Após uma apresentação geral dos dados, passamos a analisar estas 5

variáveis nesta ordem: 1) *Fenômeno*; 2) *Palavra*; 3) *Estilo*; 4) *Liderança*; e 5) *Indivíduo*. Dentro dessa análise, avaliamos as variáveis *Liderança* e *Indivíduo*, relacionando-as com *Fenômeno*, *Palavra* e *Estilo*.

Inicialmente, vimos que o resultado geral de 17,4% de cancelamento nos dados do estudo transversal aponta para a existência de variabilidade. Na avaliação da variável *Fenômeno*, constatamos que é possível que o Fenômeno 1 (r #) esteja lexicalmente mais difundido (CHEN; WANG, 1975; BYBEE, 2001) do que os Fenômenos 2 (v-v) e 3 (cc).

Na análise da variável *Palavra*, vimos que os resultados apontam para a dinamicidade e a plasticidade do léxico (OLIVEIRA, M., 1995; BYBEE, 2001; PIERREHUMBERT, 2001b), na medida em que palavras relacionadas a um mesmo fenômeno não tiveram um comportamento homogêneo. Ao contrário, vimos que observar o comportamento das palavras individualmente é importante, tendo em vista que encontramos, nelas, padrões de difusão lexical da mudança sonora (BYBEE, 2001; CHEN; WANG, 1975).

Na análise da variável *Estilo*, vimos que o caráter probabilístico da linguagem (BOD; HAY; JANNEDY, 2003) é atuante, já que há gradiência (PIERREHUMBERT, 2002) entre os diferentes estilos. Também observamos que tipos de estilo podem ser estatisticamente diferentes quanto a um fenômeno, mas não quanto a um outro. O estilo *fala espontânea*, que favorece mais fenômenos de mudança sonora, atua como um “motor” da variabilidade, por ser informal e, como propusemos, talvez pelo fato de apresentar um maior número de exemplares do que os outros estilos (PIERREHUMBERT, 2001a).

Ao avaliar a variável *Liderança* com relação a todos os 3 fenômenos ao mesmo tempo, verificamos que existe um comportamento de tendência, segundo o qual, em conjunto, os líderes se comportam como líderes, assim como também em conjunto, os não-líderes se comportam como não-líderes (LABOV, 2001, v. 2). Contudo, vimos que, ao analisar os três fenômenos separadamente, na análise da relação entre *Liderança* e *Fenômeno*, líderes e não-líderes apresentaram padrões diferentes com relação a tais fenômenos (BYBEE, 2001; SANGESTER, 2002). Isso por que, enquanto os líderes mantiveram índices de cancelamento parecidos em cada um dos fenômenos, os não-líderes apresentaram índices estatisticamente diferentes para cada um deles.

Na análise da relação entre *Liderança* e *Palavra*, vimos que líderes apresentam maior variabilidade e multiplicidade de exemplares de uma mesma palavra (BYBEE, 2001; PIERREHUMBERT, 2001b) do que não-líderes. Já na análise da relação da *Liderança* com *Estilo*, observamos que, enquanto para os líderes há uma distinção bem nítida entre estilo informal (*fala espontânea*) e estilo formal (*leitura e nomeação de figura*), para os não-líderes não há, visto que um estilo informal (*fala espontânea*) foi agrupado com um estilo formal (*nomeação de figura*) (LABOV, 1972; ABAURRE, 1981; SANGSTER, 2002). Nesse ponto, destacamos que, possivelmente, para os não-líderes, a *fala espontânea* seria um estilo mais formal do que informal (LABOV, 1972; MARCHUSCHI, 2003).

Por último, na análise da variável *Indivíduo*, destacamos a importância de se considerar o indivíduo com unidade de análise (OLIVEIRA, M., 1992; SANGESTER, 2002), tendo em vista que encontramos, dentro das categorias *líder* e *não-líder*, variabilidade individual significativa. Na análise da relação entre *Indivíduo* e *Fenômeno*, vimos que os indivíduos lidam com diferentes fenômenos de forma probabilística (BOD; HAY;

JANNEDY, 2003). Ressaltamos a importância de se avaliar fenômenos diferentes para os indivíduos (SANGSTER, 2002), a fim de se observarem variações inter e intra-individuais. Na análise da relação *Indivíduo e Palavra*, observamos que a organização do componente fonológico varia individualmente em função da *Palavra* e que indivíduos diferentes apresentam léxicos e gramáticas diferentes (HOPPER, 1987; BYBEE, 2001; REINECKE, 2006). Na análise da relação *Indivíduo e Estilo*, destacamos, mais uma vez, que a organização do componente fonológico pode variar, individualmente, de acordo com o estilo (SANGSTER, 2002).

No próximo capítulo, fazemos análises de caráter longitudinal, abordando, em grande parte, o estudo longitudinal que realizamos com 2 dos 12 indivíduos desta pesquisa.

## CAPÍTULO 6

### ESTUDO LONGITUDINAL: ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

#### 6.1. Introdução

Este capítulo é dividido em duas partes. Na primeira, que ocupa a seção 6.2, realizamos uma análise, na qual avaliamos se houve mudança estatisticamente significativa, para os 12 indivíduos em estudo, do corpus de origem de cada um deles (corpora 1, 2 e 3) para o corpus atual. Essa análise comparativa tem um caráter longitudinal, visto que observa a produção lingüística dos mesmos indivíduos em tempos diferentes: em pesquisas anteriores e na atual.

Na segunda parte deste capítulo, a qual é maior e ocupa a seção 6.3, abordamos o estudo longitudinal aqui realizado com 2 dos 12 indivíduos desta pesquisa, apresentando e analisando os dados e discutindo os resultados de tal estudo. Inclusive, comparamos a produção lingüística desses 2 indivíduos no estudo longitudinal e no estudo transversal, o que se configura como uma análise de cunho longitudinal também (Cf. seção 6.3.2.1.1). A seção 6.4 sumariza este capítulo.

## 6.2 Análise comparativa do comportamento dos 12 indivíduos no seu corpus de origem e na pesquisa atual

Nesta pesquisa, uma de nossas propostas (Cf. seção 4.2 do capítulo 4) é comparar o comportamento lingüístico dos 12 indivíduos aqui analisados no seu corpus de origem (corpora 1, 2 ou 3) e na presente pesquisa. Como foi explicitado (Cf. seção 4.3 do capítulo 4), esses 12 indivíduos foram selecionados de três pesquisas anteriores, cada uma focalizando um fenômeno específico. Para lembrar de que corpora/ pesquisa cada indivíduo é proveniente e em qual fenômeno cada indivíduo foi avaliado nessas pesquisas anteriores, apresentamos o QUADRO 10, no qual os líderes estão destacados em cinza e os não-líderes não apresentam qualquer destaque:

**QUADRO 10: Informações sobre corpus/ pesquisa e fenômeno para os 12 indivíduos**

<b>Indivíduo</b>	<b>Corpus/ pesquisa</b>	<b>Fenômeno</b>
Indivíduo 1 – L1W	Corpus 1/ Huback (2003)	1. Cancelamento de “r” final em nominais
Indivíduo 2 – L1V	Corpus 1/ Huback (2003)	1. Cancelamento de “r” final em nominais
Indivíduo 3 – N1A	Corpus 1/ Huback (2003)	1. Cancelamento de “r” final em nominais
Indivíduo 4 – N1L	Corpus 1/ Huback (2003)	1. Cancelamento de “r” final em nominais
Indivíduo 5 – L2L	Corpus 2/ Fontes Martins (2001)	2. Cancelamento de líquida intervocálica
Indivíduo 6 – L2R	Corpus 2/ Fontes Martins (2001)	2. Cancelamento de líquida intervocálica
Indivíduo 7 – N2L	Corpus 2/ Fontes Martins (2001)	2. Cancelamento de líquida intervocálica
Indivíduo 8 – N2M	Corpus 2/ Fontes Martins (2001)	2. Cancelamento de líquida intervocálica
Indivíduo 9 – L3R	Corpus 3/ Cristóforo Silva (2003)	3. Cancelamento de encontro consonantal tautossilábico
Indivíduo 10 – L3J	Corpus 3/ Cristóforo Silva (2003)	3. Cancelamento de encontro consonantal tautossilábico
Indivíduo 11 – N3Z	Corpus 3/ Cristóforo Silva (2003)	3. Cancelamento de encontro consonantal tautossilábico
Indivíduo 12 – N3G	Corpus 3/ Cristóforo Silva (2003)	3. Cancelamento de encontro consonantal tautossilábico

Na primeira coluna do QUADRO 10, os 12 participantes deste estudo são apresentados, cada qual com o seu código específico (Cf. seção 4.3 do capítulo 4). A segunda coluna

discrimina o corpus/ pesquisa anterior de que o informante foi selecionado. Já a terceira coluna mostra o fenômeno em que cada indivíduo foi analisado nessa pesquisa anterior.

Assim, por exemplo, o Indivíduo 1 – L1W é proveniente do corpus 1, da pesquisa de Huback (2003), sendo que ele foi avaliado somente quanto ao Fenômeno 1 (r #) em tal pesquisa. Dessa forma, esse indivíduo tem comparadas, na pesquisa de Huback (2003) e na presente pesquisa, somente as suas produções relativas ao Fenômeno 1 (r #), já que L1W não foi avaliado quanto aos outros dois fenômenos em estudo na pesquisa de Huback (2003). O mesmo ocorre com os outros 11 indivíduos: neste momento, eles só podem ser avaliados quanto a um fenômeno, que é aquele no qual eles são avaliados, simultaneamente, na pesquisa anterior de que fazem parte e na pesquisa atual.

Na TAB. 18, apresentamos, para cada um dos 12 indivíduos (6 líderes, destacados em cinza; 6 não-líderes, sem destaque), os percentuais de cancelamento no corpus de que cada um deles é proveniente e no corpus atual:

**TABELA 18: O comportamento dos 12 indivíduos no seu corpus de origem e no corpus atual**

Fenômeno 1 (r #)	Corpus 1- Huback (2003)		Corpus atual	
	n e % de cancelamento	total de ocorrências	n e % de cancelamento	total de ocorrências
Indivíduo 1 – L1W	23 (74%)	31	42 (74%)	57
Indivíduo 2 – L1V	13 (68%)	19	5 (15%)	33
Indivíduo 3 – N1A	4 (6%)	65	7 (28%)	25
Indivíduo 4 – N1L	2 (3%)	60	12 (20%)	60
Fenômeno 2 (v-v)	Corpus 1- Fontes Martins (2001)		Corpus atual	
	n e % de cancelamento	total de ocorrências	n e % de cancelamento	total de ocorrências
Indivíduo 5 – L2L	106 (19%)	550	73 (31%)	237
Indivíduo 6 – L2R	41 (11%)	365	25 (33%)	75
Indivíduo 7 – N2L	0 (0%)	431	2 (1%)	233
Indivíduo 8 – N2M	5 (1%)	430	5 (2%)	223
Fenômeno 3 (cc)	Corpus 1- Cristóvão Silva (2003)		Corpus atual	
	n e % de cancelamento	total de ocorrências	n e % de cancelamento	total de ocorrências
Indivíduo 9 – L3R	83 (54%)	153	12 (16%)	77
Indivíduo 10 – L3J	21 (43%)	49	18 (51%)	35
Indivíduo 11 – N3Z	0 (0%)	41	2 (4%)	50
Indivíduo 12 – N3G	2 (3%)	61	10 (28%)	36

Na TAB. 18, para cada informante (1ª coluna), é discriminado o número e o percentual de cancelamento em relação ao total de ocorrências produzidas por ele no seu corpus de origem (2ª e 3ª colunas) e no corpus atual (4ª e 5ª colunas). Como se pode observar pela análise dessa tabela, há indivíduos que apresentam um percentual de cancelamento igual ou muito parecido nos corpora anterior e atual. Esse é o caso, por exemplo, do Indivíduo 1 – L1W que apresenta exatamente o mesmo percentual de cancelamento no seu corpus de origem (Corpus 1 – HUBACK, 2003) e no corpus atual, quanto ao Fenômeno 1 (r #) em que é analisado: 74% de cancelamento de “r” final em nominais.

Por outro lado, analisando a TAB. 18, percebemos que há indivíduos que apresentaram diferenças percentuais de cancelamento quanto aos corpora anterior e atual. Como exemplo, podemos citar o Indivíduo 2 – L1V, o qual, assim como L1W, é líder e proveniente do Corpus 1, sendo também avaliado quanto ao Fenômeno 1 (r #). Contudo,

L1V que apresentou um percentual de 68% de cancelamento no seu corpus de origem – o que é muito próximo do L1W –, apresenta um percentual consideravelmente menor (4,5 vezes menor) no corpus atual: apenas 15% de cancelamento. Tal resultado relativo ao Indivíduo 2 – L1V aponta para o fato de que é possível que essa variação intra-individual seja relevante.

A fim de conferir se as mudanças no comportamento dos indivíduos, observadas na TAB. 18, foram significativas, realizamos um teste de proporção<sup>53</sup>, o qual pode ser conferido na TAB. 19. Mais especificamente, esse teste verificou se houve mudanças significativas entre os percentuais de cancelamento no corpus de origem de cada informante e no corpus atual.

**TABELA 19: Resultados do teste de proporção**

<b>Fenômeno 1 (r #)</b>	<b>Resultado do teste de proporção</b>
Indivíduo 1 – L1W	0,958
Indivíduo 2 – L1V	0,00
Indivíduo 3 – N1Ab	0,00
Indivíduo 4 – N1L	0,02
<b>Fenômeno 2 (v-v)</b>	<b>Resultado do teste de proporção</b>
Indivíduo 5 – L2L	0,00
Indivíduo 6 – L2R	0,00
Indivíduo 7 – N2L	0,15
Indivíduo 8 – N2M	0,33
<b>Fenômeno 3 (cc)</b>	<b>Resultado do teste de proporção</b>
Indivíduo 9 – L3R	0,00
Indivíduo 10 – L3J	0,437
Indivíduo 11 – N3Z	0,15
Indivíduo 12 – N3G	0,002

<sup>53</sup> Esse teste foi realizado com o programa estatístico Minitab for Windows, versão 13. Nessa análise, os casos em que a diferença foi considerada estatisticamente significativa apresentam P (valor na 2ª coluna da TAB. 19) menor que 0,05.

A TAB. 19 apresenta os 12 indivíduos desta pesquisa (6 líderes, destacados em cinza; 6 não-líderes, sem destaque) e os fenômenos em que cada um deles foi avaliado na 1ª coluna. Na 2ª coluna dessa tabela, vemos o resultado do teste de proporção realizado.

Como se pode perceber, esse teste confirma que, para o Indivíduo 1 – L1W, não foi estatisticamente significativa a diferença entre os percentuais de cancelamento dele no Corpus 1, de que é proveniente, e no corpus atual (Cf. valor de 0,958 na 2ª linha da TAB. 19). Esse teste confirma ainda que, para o Indivíduo 2 – L1V, ao contrário, foi estatisticamente significativa a diferença entre os percentuais de cancelamento dele no Corpus 1, de que também é proveniente, e no corpus atual (Cf. valor de 0,00 na 3ª linha da TAB. 19). A mudança no comportamento de L1V e a estabilidade no de L1W são referidas na já apresentada afirmação de Paiva e Duarte (2003, p. 17): “[a] comparação de amostras de fala dos mesmos falantes em diferentes pontos do tempo, permite captar mudanças ou estabilidade no comportamento lingüístico do indivíduo [...]”. Assim, esse resultado demonstra a importância de se realizar um estudo longitudinal para avaliar a variabilidade individual.

Ainda com relação à TAB. 19, podemos notar que, para 5 indivíduos (2 líderes – L1W, L3J – e 3 não-líderes – N2L, N2M, N3Z), não foi estatisticamente significativa a diferença de percentual de cancelamento nos corpora anterior e atual. Contudo, para a maioria dos indivíduos, 7 deles (sendo 4 líderes – L1V, L2L, L2R, L3R – e 3 não-líderes – N1A, N1L, N3G), foi estatisticamente significativa essa diferença.

O mais importante a ser destacado da análise feita nesta seção é que houve variabilidade intra-individual (SANGSTER, 2002) significativa com relação ao tempo que se passou das coletas das pesquisas anteriores (uma de 2001 e duas de 2003, ou seja, há um

intervalo de 3 a 5 anos das pesquisas anteriores para a atual) para a coleta da pesquisa atual, em relação a alguns indivíduos, mais especificamente, a 7 indivíduos. Vale destacar que, na próxima seção, que apresenta o estudo longitudinal realizado neste trabalho, novamente, poderemos avaliar o fator tempo implícito nesse tipo de estudo. Contudo, o tempo observado foi menor: de 6 meses apenas.

Como discutimos na seção 2.4 do capítulo 2, Paiva e Duarte (2003) também apontam variações no comportamento lingüístico de indivíduos específicos em função do tempo. Razões para essa alteração no comportamento dos indivíduos, no caso os 7 citados (L1V, L2L, L2R, L3R, N1A, N1L, N3G), merecem maior investigação; contudo, aqui levantamos a possibilidade de aspectos extralingüísticos estarem ligados à estabilidade no comportamento lingüístico do Indivíduo 1 – L1W e à instabilidade no do Indivíduo 2 – L1V. Do intervalo da pesquisa de Huback (2003) – em que os Indivíduos 1 e 2 foram avaliados pela primeira vez somente quanto ao Fenômeno 1 (r #) – para a pesquisa atual (com um intervalo de 3 anos aproximadamente), enquanto o Indivíduo 1 não passou por nenhuma mudança significativa em sua vida, ao contrário, o Indivíduo 2 passou por muitas mudanças importantes, inclusive, de escolaridade. Paiva e Duarte (2003) apontam que mudanças na vida do indivíduo (mudança de escolaridade, de emprego, o fato de casar, ter filhos, etc.) podem contribuir para mudanças no comportamento lingüístico deste.

Esses dois indivíduos, parecidos em muitos aspectos<sup>54</sup>, apresentam, contudo, personalidades e perspectivas de vida muito diferentes. O Indivíduo 1 se mostrou

---

<sup>54</sup> Ambos os Indivíduos 1 e 2, além de serem líderes provenientes do Corpus 1, são homens, de classe baixa, moram na periferia de Belo Horizonte e trabalham no mesmo local exercendo a mesma função. Há aproximadamente 3 anos, na coleta da pesquisa anterior (HUBACK, 2003) de que fizeram parte, esses dois indivíduos tinham a mesma escolaridade (ensino fundamental). Contudo, na coleta atual, enquanto o Indivíduo 1 permaneceu com essa escolaridade, o Indivíduo 2 aumentou o seu grau de escolarização, concluindo o 2º grau e até fazendo um curso técnico de elétrica na Escola de Engenharia da UFMG.

pacato, acomodado, dizendo-se, inclusive, totalmente satisfeito com a vida que tem, não querendo alterar nada nela. Já o Indivíduo 2 se mostrou uma pessoa dinâmica, com ambição, planos para o futuro e intenção de obter ascensão social, motivo pelo qual aumentou o seu grau de escolaridade: concluiu o 2º grau e fez um curso técnico de elétrica. Sangster (2002), como se mencionou na seção 2.3.2 do capítulo 2, encontrou, em sua pesquisa, que falantes com maior ambição apresentaram maior variação quanto aos fenômenos por ela estudados. Isso se relaciona com esse nosso achado para o Indivíduo 2 que, quanto ao Fenômeno 1 (r #), apresentou variabilidade estatisticamente significativa da pesquisa anterior de que fez parte (HUBACK, 2003) para a pesquisa atual. Podemos afirmar que essa variabilidade do Indivíduo 2 aponta para a plasticidade da gramática (BYBEE, 2001).

Tendo em vista até mesmo a limitação de tempo deste trabalho, não nos é possível investigar, detalhadamente, os motivos pelos quais ocorre (ou não) variabilidade no comportamento de todos os indivíduos em análise. Seria interessante que trabalhos futuros se ocupassem dessa tarefa. Inclusive, na análise desta seção, não pudemos avaliar os 12 indivíduos quanto a todos os 3 fenômenos em estudo; eles foram avaliados somente quanto a um fenômeno, o que explicamos anteriormente (Cf. QUADRO 9). Portanto, uma pesquisa posterior relevante seria investigar esses indivíduos quanto a mais de um fenômeno, em um intervalo de tempo maior, como feito na análise desta seção.

Passemos, neste momento, à outra parte deste capítulo: a análise dos dados do estudo longitudinal aqui realizado com 2 dos 12 indivíduos do estudo transversal, o que é feito na seção 6.3, a seguir.

### 6.3 Análise e discussão dos resultados do banco de dados do estudo longitudinal

#### 6.3.1 Visão geral dos dados do estudo longitudinal

Como foi apontado no capítulo 4, “Metodologia”, os dados do estudo longitudinal aqui realizado são relativos à produção de 2 dos 12 informantes do corpus do estudo transversal: uma informante líder, Indivíduo 6 – L2R, e um informante não-líder, Indivíduo 11 – N3Z (Cf. QUADRO 6 da seção 4.3). Como já se mencionou na seção 4.5.2.1, esses informantes foram escolhidos por terem se mostrado mais acessíveis, por terem aceitado colaborar com esta pesquisa prontamente.

O banco de dados do estudo longitudinal reflete a produção lingüística dos indivíduos L2R e N3Z. Tal banco conta com um total de 6.598 ocorrências para 1.646 tipos de palavras que apresentam contexto fonético para a realização dos 3 fenômenos em análise. Desse total de 6.598 ocorrências, 6.012 ocorrências (91,1%) não apresentaram cancelamento dos sons observados, enquanto que 586 ocorrências (8,9%) apresentaram, o que podemos ver pela TAB. 20:

**TABELA 20: Dados do estudo longitudinal**

Total de tipos de palavras	Total de ocorrências	Ocorrências que apresentaram cancelamento (n/ %)	Ocorrências que não apresentaram cancelamento (n/ %)
1.646	6.598	586 (8,9%)	6.012 (91,1%)

É interessante comparar os dados do estudo longitudinal com os do transversal e ver que o número total de ocorrências é superior no estudo longitudinal: enquanto há 4.955 dados no estudo transversal (Cf. TAB. 7 da seção 5.2.1), há 6.598 no longitudinal, o que totaliza 1.643 dados ou 25% de dados a mais no longitudinal. O estudo transversal conta com um tempo de gravação superior ao longitudinal: enquanto o transversal soma, aproximadamente, 13 horas de gravação, o longitudinal soma em torno de 6 horas (Cf. capítulo 4). Contudo, o estudo transversal apresentou menos dados, porque, vale lembrar, nele, foram consideradas somente aquelas 30 palavras referentes aos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em análise (Cf. QUADRO 8 da seção 4.4.3.4). Já no estudo longitudinal, como se disse ainda nesta seção, foram consideradas todas as palavras que apresentavam contexto para os referidos fenômenos.

Outra observação a ser feita é que o percentual de cancelamento no estudo transversal foi maior: enquanto no estudo longitudinal, 8,9% das palavras apresentaram cancelamento, no estudo transversal, 17,4% (Cf. TAB. 7 da seção 5.2.1) das palavras apresentaram; ou seja, no estudo transversal, houve praticamente o dobro de cancelamento. É possível que o Indivíduo não-líder – N3Z, avaliado no estudo longitudinal, tenha contribuído para abaixar o percentual de cancelamento de tal estudo, visto que, dentre todos os indivíduos analisados nesta pesquisa, ele foi o que apresentou o menor índice de cancelamento (Cf. TAB. 16 da seção 5.2.3.2), com 0,9%. N3Z até diminuiu um pouco o seu percentual de cancelamento no estudo longitudinal, passando para 0,7% (Cf. TAB. 21 da próxima seção). De toda maneira, o próprio percentual de 15,7% de cancelamento do Indivíduo líder – L2R (Cf. TAB. 21 da próxima seção) também contribuiu para esse baixo índice de 8,9% de cancelamento do estudo longitudinal.

Além da variável dependente *Cancelamento* (que apresenta duas categorias: 1. cancela, 2. não-cancela), há outras 3 variáveis independentes que foram aqui consideradas:

- a) *Indivíduo* → que apresenta duas categorias: 1. Indivíduo – L2R, 2. Indivíduo – N3Z  
(Cf. QUADRO 6 da seção 4.3);
- b) *Coleta* → que apresenta 6 categorias: Coletas 1, 2, 3, 4, 5 e 6;
- c) *Palavra* → esta variável apresentava inicialmente 1.646 categorias correspondentes a 1.646 tipos de palavra. Como será explicitado na seção 6.3.2.3, a fim de tornar exequível a análise desta variável, fizemos um recorte, de modo que chegamos a apenas 16 categorias/ tipos de palavra.

Nas próximas seções (de 6.3.2. a 6.3.2.3), analisaremos, separadamente, cada uma dessas 3 variáveis independentes consideradas no estudo longitudinal.

### **6.3.2 Análise das variáveis independentes *Indivíduo, Coleta, Palavra***

#### **6.3.2.1 Análise da variável *Indivíduo***

Vejamos a TAB. 21 que apresenta o percentual de cancelamento e o peso relativo das categorias da variável *Indivíduo*, Indivíduos N3Z e L2R:

**TABELA 21: Percentual de cancelamento e peso relativo dos indivíduos do estudo longitudinal**

	Indivíduos		Total
	N3Z	L2R	
Não-Cancelamento	2968 99,3%	3044 84,3%	6012 91,1%
Cancelamento	21 0,7%	565 15,7%	586 8,9%
Total	2989	3609	6598
Peso relativo	.04	.96	

Chi-Square Tests: 0,000

Como se pode perceber pela análise da TAB. 21, o não-líder – N3Z apresenta um percentual de cancelamento muito baixo (0,7%). Fazendo uma analogia com os efeitos de protótipo de Pierrehumbert (2001b), é como se N3Z fosse o membro central da categoria “não-líder”, já que ele tem um comportamento bem típico de não-líder, apresentando o menor percentual de cancelamento dentre todos os informantes desta pesquisa, tanto no corpus do estudo transversal, quanto no do longitudinal. Já a líder L2R apresenta um percentual de cancelamento maior (15,7%), podendo ser considerada líder. Contudo, ela não chega a ser o membro central da categoria “líder”, visto que há outros líderes no estudo transversal que apresentam percentuais de cancelamento bem superiores ao dela (Cf. na TAB. 16 da seção 5.2.3.2, por exemplo, os indivíduos L1W com 42,9% de cancelamento e L2L com 38,7%).

O valor 0,000 do teste de Chi-square, abaixo da TAB. 21, mostra que a diferença nos percentuais de cancelamento dos indivíduos NZ3 e L2R é estatisticamente significativa. Ainda, como se pode notar pela análise da TAB. 21, esse resultado é confirmado pelos valores de peso relativo: enquanto L2R favorece grandemente o cancelamento com .96, ao contrário, N3Z desfavorece muito com .04. Isso aponta para o fato de que L2R que é líder se comporta como líder realmente, assim como N3Z que é não-líder se comporta

como não-líder, nas 6 coletas realizadas no estudo longitudinal (LABOV, 2001, v. 2). Apesar de que L2R não poder ser considerada como o protótipo da sua categoria “líder”, ao contrário de N3Z que seria o protótipo de sua categoria “não-líder” (PIERREHUMBERT, 2001b). Esse fato, inclusive, aponta para a distribuição probabilística das categorias “líder” e “não-líder” em um contínuo gradual (PIERREHUMBERT, 2002; BOD; HAY; JANNEDY, 2003), de modo que na representação da TAB. 9, enquanto N3Z estaria à esquerda da reta, bem próximo de 0 (não-líder), L2R estaria à direita, mas não tão próxima de 1 (líder).

Dentro do estudo longitudinal, há uma observação muito interessante a ser feita a respeito do léxico de L2R e N3Z. Esses indivíduos tiveram um tempo de gravação muito parecido no estudo longitudinal: aproximadamente, 3 horas de gravação cada um. Contudo, houve uma diferença considerável entre eles quanto ao número de tipos de palavra e de ocorrências, ou seja, quanto às frequências de tipo e de ocorrência.

A líder L2R apresentou 3.609 ocorrências<sup>55</sup> para 818 tipos de palavras, o que significa uma média de 4,4 ocorrências por tipo de palavra. Já o não-líder N3Z apresentou 2.989 ocorrências para 1090 tipos de palavras, o que significa uma média de 2,7 ocorrências por tipo de palavra. Esse resultado pode ser indício de que o indivíduo N3Z, que apresenta mais tipos de palavras, com 280 (26%) tipos de palavras a mais do que a líder L2R, faz uso de um léxico mais amplo do que esta informante. Tal fato aponta a possibilidade de indivíduos diferentes apresentarem, no uso, léxicos (e conseqüentemente, gramáticas) diferentes (BYBEE, 2001; REINECKE, 2006), tanto em termos de frequência de tipo quanto de ocorrência (BYBEE, 2001).

---

<sup>55</sup> Deve-se ressaltar que esses números de tipo e de ocorrência de palavra para os Indivíduos L2R e N3Z se referem somente às palavras que apresentaram contexto para a realização dos Fenômenos 1 (r #), 2 (v-v) e 3 (cc) em análise.

Levantamos aqui duas razões possíveis para o Indivíduo N3Z poder ter, no uso, um léxico mais amplo, com mais tipos de palavras, do que L2R. A primeira é que N3Z é mais velho, tendo 56/ 57 anos de idade quando das coletas, do que a líder L2R, que tinha 26 anos de idade quando das coletas. Por ele ser mais velho, é provável que ele tenha tido uma exposição maior ao léxico da língua, ou seja, mais experiência lingüística (BYBEE, 2001). Com isso, é possível que ele tenha adquirido maior número de categorias (PIERREHUMBERT, 2001b) do que L2R.

Ainda, a segunda razão para o não-líder N3Z poder ter um léxico mais amplo que líder L2R está relacionada ao fato de que, apesar de N3Z não ter curso superior completo, ele é muito culto, erudito e bastante sensível à linguagem. Ele, inclusive, escreve poemas, contos, tendo até um livro já escrito que deseja publicar. O Indivíduo N3Z passa grande parte de seu tempo livre lendo livros, assistindo a filmes, tocando e ouvindo música. Já a informante L2R, que tem curso superior completo, é uma pessoa instruída e gosta de executar tarefas práticas. Ela, que é contadora, em seu horário livre, não se dedica tanto à leitura, por exemplo. Ela prefere sair com os amigos, cozinhar, conversar com os seus familiares, o que ela própria relatou nas coletas.

A sensibilidade ou consciência maior do não-líder N3Z para a linguagem faz com que ele apresente uma fala cuidada, quase que constantemente tendo uma elocução formal (LABOV, 1972), muito próxima da norma culta. Isso o leva a apresentar pouca variabilidade em sua fala, dito de outra forma, uma menor multiplicidade de exemplares de uma categoria ou menos exemplares por tipo de palavra (PIERREHUMBERT, 2001b). Como consequência, tal indivíduo é menos suscetível a realizar fenômenos de mudança e variação sonora.

Outro fato interessante relacionado ao léxico desses indivíduos é que L2R que apresentou menos tipos de palavra, por outro lado, apresentou mais ocorrências: 620 (17%) ocorrências a mais do que N3Z. Esse fato pode estar relacionado a uma questão de taxa de elocução (BARBOSA, 2000). É possível que, enquanto L2R apresenta uma taxa de elocução mais rápida, N3Z apresenta uma taxa mais lenta, contudo, não chegamos a realizar medições a respeito desse parâmetro prosódico. Seria, interessante que trabalhos futuros avaliassem se existe uma correlação entre taxa de elocução e variabilidade individual, e até mesmo entre taxa de elocução e a questão da liderança, investigando se líderes apresentam uma taxa de elocução mais rápida do que não-líderes, como ocorreu com os indivíduos L2R e N3Z.

Na próxima seção, esses dois indivíduos são comparados nos bancos de dados transversal e longitudinal, de forma que tal análise apresenta um caráter longitudinal também.

#### **6.3.2.1.1 Comparação dos indivíduos L2R e N3Z nos dois bancos de dados: transversal e longitudinal**

Tendo em vista que os Indivíduos L2R e N3Z foram os únicos que participaram, simultaneamente, dos dois estudos desta pesquisa, o transversal e o longitudinal, fazemos uma comparação desses indivíduos nesses dois estudos. Assim, a TAB. 22 nos mostra essa análise:

**TABELA 22: Indivíduos L2R e N3Z nos estudos transversal e longitudinal**

		<b>Informante líder L2R</b>	<b>Informante não- líder N3Z</b>
<b>Banco de dados estudo transversal</b>	<b>% e n cancelamento</b>	20% (45/ 225)	0,9% (3/ 321)
	<b>peso relativo</b>	.60	.13
<b>Banco de dados estudo longitudinal</b>	<b>% e n cancelamento</b>	15,7% (565/3.609)	0,7% (21/ 2.989)
	<b>peso relativo</b>	.96	.04

Pela TAB. 22, vemos que a líder L2R, no estudo transversal, apresentou 20% de cancelamento dos dados, o que corresponde a 45 palavras de um total de 225 ocorrências, e teve peso relativo de .60. No estudo longitudinal, L2R apresentou 15,7% de cancelamento dos dados, o que corresponde a 565 palavras de um total de 2.989 ocorrências, e o seu peso relativo foi de .96.

Já quanto ao não-líder N3Z, vemos que, no estudo transversal, ele apresentou 0,9% de cancelamento dos dados, o que corresponde a 3 palavras de um total de 321 ocorrências, sendo que o seu peso relativo foi de .13. No estudo longitudinal, N3Z apresentou 0,7% de cancelamento dos dados, o que corresponde a 21 palavras de um total de 3.609 ocorrências, sendo que o seu peso relativo foi de .04.

Ao analisar a TAB. 22, vemos que não houve muita diferença no comportamento dos indivíduos L2R e N3Z quanto aos estudos transversal e longitudinal, quando considerados os 3 fenômenos de cancelamento aqui avaliados. Apesar de, nesses estudos, a diferença de peso relativo ser considerável para os dois indivíduos, a diferença percentual é relativamente pequena.

Montamos um novo banco de dados considerando somente os dados dos Indivíduos L2R e N3Z nos estudos transversal e longitudinal, a fim de verificar se houve diferença estatisticamente significativa no comportamento deles quanto aos dois estudos. Assim,

realizamos análises de contingência, nas quais foi aplicado o teste de Chi-square, para esses dois indivíduos em separado. Vejamos, então, a análise feita para L2R na TAB. 23:

**TABELA 23: Tabela de contingência entre as variáveis *Cancelamento e Informante*: somente para o Indivíduo L2R**

	Estudo		Total
	Transversal	Longitudinal	
Não-Cancelamento	180 80,0%	3044 84,1%	3224 84,1%
Cancelamento	45 20,0%	565 15,7%	610 15,9%
Total	225	3609	3834

Chi-Square Tests: 0,084

Como podemos ver pela TAB. 23, o resultado do teste de Chi-square (0,84) confirma que, realmente, não é estatisticamente significativa a diferença de cancelamento, para a informante L2R, nos estudos transversal e longitudinal. Vejamos agora a análise feita para o Indivíduo N3G na TAB. 24:

**TABELA 24: Tabela de contingência entre as variáveis *Cancelamento e Informante*: somente para o Indivíduo N3Z**

	Estudo		Total
	Transversal	Longitudinal	
Não-Cancelamento	318 99,1%	2968 99,3%	3286 99,3%
Cancelamento	3 0,9%	21 0,7%	24 0,7%
Total	321	2989	3310

Chi-Square Tests: 0,642

A TAB. 24 nos mostra que, para o indivíduo N3Z, a diferença de cancelamento nos estudos transversal e longitudinal é menos significativa ainda do que para L2R, o que podemos perceber pelo resultado do Chi-Square, 0,642. Com isso, podemos concluir que esses dois indivíduos não alteraram o seu comportamento, de forma estatisticamente

significante, quanto aos dois diferentes estudos aqui realizados: o transversal e o longitudinal. É possível que, para esses indivíduos que são adultos, o período de tempo em que eles foram observados nesses dois estudos (6 meses) tenha sido pequeno, não sendo suficiente para eles apresentarem variação significativa quanto aos fenômenos em análise, até mesmo porque eles não sofreram mudanças de vida (PAIVA; DUARTE, 2003; SANGESTER, 2002). Talvez seja preciso um tempo maior para que variações individuais sejam observadas (PAIVA; DUARTE, 2003), especialmente na linguagem adulta que apresenta maior estabilidade<sup>56</sup>. Na próxima seção, com a análise variável *Coleta*, poderemos verificar essa hipótese novamente, visto que tal variável, importante no estudo longitudinal, também reflete o fator tempo, a partir do qual podem ser observadas variações no comportamento de indivíduos (PAIVA; DUARTE, 2003).

### 6.3.2.2 Análise da variável *Coleta*

Em relação à variável *Coleta*, fazemos análises em separado para os Indivíduos N3Z e L2R, a fim de avaliar o possível efeito de tal variável para cada indivíduo em particular. Assim, vejamos, primeiramente, os valores de cancelamento em cada uma das 6 coletas para o Indivíduo N3Z na TAB. 25:

---

<sup>56</sup> Crianças em fase de aquisição apresentam uma linguagem mais instável, de modo que, a depender da idade, no período de 1 mês, observam-se mudanças significativas na sua linguagem (MIRANDA, 2007; OLIVEIRA GUIMARÃES, em andamento). Ao contrário, adultos parecem têm uma linguagem mais estável, de forma que, para se notarem mudanças significativas nesta, é necessário observar um período maior de tempo, não de meses, mas de anos e até de décadas (PAIVA; DUARTE, 2003).

**TABELA 25: Tabela de contingência entre as variáveis *Cancelamento* e *Coleta*: todas as 6 coletas, somente para o não-líder N3Z**

Coleta	Cancelamento				Total
	Não-cancelamento		Cancelamento		
	n	%	n	%	
Coleta 1	470	98,9%	5	1,1%	475
Coleta 2	515	99,6%	2	0,4%	517
Coleta 3	552	99,3%	4	0,7%	556
Coleta 4	400	99,8%	1	0,2%	401
Coleta 5	547	98,8%	1	0,2%	548
Coleta 6	484	99,4%	8	1,6%	492
Total	2968	99,3%	21	0,7%	2989

Chi-Square Tests: 0,054

Como se pode perceber pela análise da TAB. 25, para o Indivíduo N3Z, não foi estatisticamente significativa a diferença entre as 6 coletas, apesar de o valor do Chi-square de 0,054 estar bem próximo do limite de significância de 0,05. Ou seja, N3Z não apresentou variação relevante quanto aos três fenômenos em estudo, considerando-se as 6 coletas aqui realizadas. Vejamos agora a análise feita para a líder – L2R na TAB. 26:

**TABELA 26: Tabela de contingência entre as variáveis *Cancelamento* e *Coleta*: todas as 6 coletas, somente para a líder L2R**

Coleta	Cancelamento				Total
	Não-cancelamento		Cancelamento		
	n	%	n	%	
Coleta 1	389	87,0%	58	13,0%	447
Coleta 2	735	84,8%	132	15,2%	867
Coleta 3	645	84,0%	123	16,0%	768
Coleta 4	396	87,2%	58	12,8%	454
Coleta 5	408	86,1%	66	13,9%	474
Coleta 6	471	78,6%	128	21,4%	599
Total	3044	84,3%	565	15,7%	3609

Chi-Square Tests: 0,001

Notamos, pelo valor do Chi-square (0,001) na TAB. 26, que é estatisticamente significativa a diferença entre as 6 coletas para a informante L2R, ao contrário do que ocorreu com o informante N3Z. Contudo, para L2R, vemos que essa diferença estatisticamente significativa entre as 6 coletas é, na verdade, promovida pela Coleta 6.

Quando retiramos a Coleta 6, vemos, pelo valor do Chi-square (0,428) abaixo da TAB. 27, que não é estatisticamente significativa a diferença entre as demais coletas:

**TABELA 27: Tabela de contingência entre as variáveis *Cancelamento* e *Coleta*: sem a Coleta 6, somente para a líder L2R**

Coleta	Cancelamento				Total
	Não-cancelamento		Cancelamento		
	n	%	n	%	
Coleta 1	389	87,0%	58	13,0%	447
Coleta 2	735	84,8%	132	15,2%	867
Coleta 3	645	84,0%	123	16,0%	768
Coleta 4	396	87,2%	58	12,8%	454
Coleta 5	408	86,1%	66	13,9%	474
Total	2573	85,5%	437	14,5%	3010

Chi-Square Tests: 0,428

Na TAB. 27, ao desconsiderarmos a Coleta 6 para L2R, vemos que era tal coleta que estava tornando estatisticamente significativa a diferença entre as seis coletas (TAB. 27). Dessa forma, percebemos que, assim como para N3Z, não há diferença significativa entre as seis coletas para a informante L2R. Com isso, podemos concluir que, na verdade, o fator tempo (PAIVA; DUARTE, 2003), implícito nas seis diferentes coletas, não foi significativo em nossa análise.

Possivelmente, isso ocorreu pelo fato de o intervalo entre uma coleta e outra (um mês em média, num total de 6 meses de coleta) ter sido pequeno para se observarem variações no comportamento dos Indivíduos N3Z e L3R, quanto aos três fenômenos em análise. Como se destacou na seção anterior, talvez, com um intervalo de tempo maior, poderia ser observado variabilidade intra-individual significativa (PAIVA; DUARTE, 2003; SANGSTER, 2002), como ocorreu com alguns indivíduos deste estudo, dentre eles, a líder L2R, na análise da seção 6.2, em que houve um intervalo de 3 a 5 anos de uma coleta para a outra.

Vale mencionar que, nesse pequeno intervalo de 6 meses, não ocorreram mudanças significativas na vida dos Indivíduos N3Z e L3R, como ocorreu, por exemplo, com o Indivíduo 2 – L1V no intervalo de 3 anos, em que, como vimos (Cf. seção 6.2), aumentou a sua escolaridade consideravelmente (PAIVA; DUARTE, 2003).

A seguir, tratamos da última variável considerada no estudo longitudinal, a variável *Palavra* que também foi avaliada no estudo transversal (Cf. seção 5.2.3.4) e que é importante em trabalhos, como o nosso, que tem como base a Fonologia de Uso, proposta por Bybee (2001), a qual considera a palavra como unidade de análise (Cf. seção 3.2.2.1).

### **6.3.2.3 Análise da variável *Palavra***

A variável *Palavra* é aqui avaliada, pelo fato de adotarmos a Fonologia de Uso (BYBEE, 2001) que considera a palavra como a unidade básica de análise (Cf. seção 3.2.2.1). Deve-se observar que seria impraticável, nesta pesquisa, trabalhar com os 1.646 tipos de palavra que ocorreram no estudo longitudinal, visto que isso resultaria em 1.646 categorias para a variável *Palavra*<sup>57</sup>. Tendo em vista esse fato, optamos por fazer um recorte dessa variável, o que reduziu as 1.646 categorias/ tipos de palavra para apenas 16 categorias/ tipos de palavra. Inclusive, é preciso dizer que montamos um outro banco de dados para analisar essas 16 palavras.

---

<sup>57</sup> Vale ressaltar que, pelo menos até onde é de nosso conhecimento, o programa estatístico SPSS roda análises com até 1000 categorias para uma variável, ou seja, não foi nem possível rodar essas 1.646 palavras nesse programa, apesar de elas terem sido rotuladas na aba “variable view”, uma a uma. Quando se tenta rodar, por exemplo, uma análise de contingência, é dada a seguinte mensagem: “Warmings – The observed number of values for Palavra exceeds the CROSSTABS limit of 1000 values. [...] This command is not executed”, ou seja, em português, “Aviso – O número de categorias observado para a variável *Palavra* excede o limite da análise de contingência que é de 1000 categorias. [...] Esse comando não é executado” (tradução nossa).

Assim, na próxima seção, 6.3.2.3.1, discutimos os critérios que nortearam esse recorte, para depois, na seção 6.3.2.3.2, procedermos à análise dessas 16 palavras do estudo longitudinal, com os Indivíduos NZ3 e L2R ao mesmo tempo e em separado.

#### **6.3.2.3.1 Critérios para o recorte da variável *Palavra***

Como foi mencionado na seção anterior, no estudo longitudinal, em razão da grande quantidade de tipos de palavra coletada (1.646 tipos), foi feito um recorte inicial em que se consideraram somente palavras com 5 ou mais ocorrências para os dois informantes analisados neste estudo, L2R e N3Z, ao mesmo tempo. Tivemos esse critério de só selecionar palavras que ocorreram para os dois indivíduos simultaneamente, já que eles são os únicos avaliados no estudo longitudinal, refletindo esse banco de dados e também para podermos fazer análises comparando tais indivíduos quanto à variável *Palavra* que é considerada como unidade de análise na Fonologia de Uso (BYBEE, 2001), teoria que adotamos. Ainda, optamos pelo número mínimo de 5 ocorrências, primeiramente, pelo fato de que, com tal número, teríamos acesso às palavras mais freqüentes do estudo longitudinal.

O limite de 5 palavras nos possibilitou um bom número de tipos de palavra (39 tipos) para ser manipulado em uma análise estatística como a que é aqui realizada. Os 39 tipos de palavra a que chegamos a partir desse recorte inicial foram, em ordem alfabética:

*agora, ali, aquela, aquele, aquilo, Belo, **Brasil**, dela, dele, dentro, ela, ele, eles, era, fala, falei, falou, grande, Horizonte, lugar, maior, melhor, mora, mostrar, mulher, outra, outro, para, Paulo, pelo, por, primeiro, queria, quero, sempre, trabalhar, trem, três, trinta.*

Note-se que, dentre essas 39 palavras, há algumas, mais exatamente 16 palavras, que estão destacadas em negrito. Ao analisar as 39 palavras, encontramos que apenas 16 delas (as que estão em negrito) coincidiam com palavras do estudo transversal que contou com 30 palavras (Cf. seção 5.2.3.4). Dessa forma, fizemos um último recorte na variável *Palavra*, de modo a considerar somente essas 16 palavras, as quais são analisadas na próxima seção.

#### **6.3.2.3.2 Análise das 16 palavras consideradas no estudo longitudinal**

Nesta seção, primeiramente, avaliamos as 16 palavras consideradas no estudo longitudinal, para os Indivíduos N3Z e L2R ao mesmo tempo. Em seguida, avaliamo-las para cada um desses indivíduos em separado. Assim, vejamos a primeira análise com a TAB. 28, a qual mostra uma análise de contingência entre as variáveis *Cancelamento* e *Palavra*:

**TABELA 28: Tabela de contingência entre as variáveis *Cancelamento* e *Palavra*: as 16 palavras consideradas no estudo longitudinal**

Palavra	Cancelamento				Total
	Não-cancelamento		Cancelamento		
	n	%	n	%	
melhor	17	81,0%	4	19,0%	21
lugar	15	83,3%	3	16,7%	18
mulher	17	89,5%	2	10,5%	19
maior	12	100%	0	0%	12
eles	72	57,1%	54	42,9%	126
para	334	72,8%	125	27,2%	459
agora	79	79,0%	21	21,0%	100
ela	364	80,9%	86	19,1%	450
ele	335	86,1%	54	13,9%	389
era	147	97,4%	4	2,6%	151
sempre	9	50,0%	9	50,0%	18
outro	31	67,4%	15	32,6%	46
primeiro	15	93,8%	1	6,3%	16
Brasil	15	100,0%	0	0%	15
grande	31	100,0%	0	0%	31
três	34	100,0%	0	0%	34
Total	1527	80,2%	378	19,8%	1905

Chi-Square Tests: 0,000

Na TAB. 28, os 16 tipos de palavras considerados no estudo longitudinal estão divididos quanto aos três fenômenos em análise: as 4 primeiras palavras, destacadas em cinza claro, se referem ao Fenômeno 1 (r #); as 6 palavras sem destaque algum se referem ao Fenômeno 2 (v-v); as últimas 6 palavras, destacadas em cinza escuro, se referem ao Fenômeno 3 (cc).

Como se pode perceber pela análise da TAB. 28, a partir do recorte feito na variável *Palavra*, chegamos a um total de 1.905 palavras para os 16 tipos aqui analisados, presentes na 1ª coluna de tal tabela. Desse total, 378 palavras ou 19,8% apresentam cancelamento dos sons em análise. É preciso ressaltar que o percentual de cancelamento de 19,8% do estudo longitudinal agora se aproxima do percentual de cancelamento do estudo transversal que foi de 17,4% (Cf. TAB. 7 da seção 5.2.1).

A proximidade nos percentuais de cancelamento desses estudos pôde ser observada a partir do momento em que foi feito o recorte da variável *Palavra*, o qual fez com os bancos de dados dos estudos transversal e longitudinal se assemelhassem. O banco do estudo longitudinal passou a contar com palavras mais freqüentes somente, como o banco do estudo transversal (Cf. seção anterior). O que esse fato demonstra é que, no caso dos três fenômenos em análise, palavras mais freqüentes em conjunto se comportam de modo semelhante (BYBEE, 2001), seja em um estudo transversal ou em um longitudinal.

Uma outra observação a se fazer a respeito das 16 palavras avaliadas no estudo longitudinal é que, pela TAB. 28, notamos que a maioria delas, mais exatamente 13, são dissílabas: *Brasil, ela, ele, eles, era, grande, lugar, maior, melhor, mulher, outro, para, sempre*. Há somente 2 palavras trissílabas (*agora, primeiro*), 1 palavra monossílaba (*três*) e nenhuma palavra polissílaba. Vale dizer que palavras dissílabas e trissílabas são, em termos de tipo (BYBEE, 2001), mais freqüentes do que palavras polissílabas e monossílabas em nossa língua (PROJETO ASPA).

Assim como nas 30 palavras do estudo transversal, nessas 16 palavras do estudo longitudinal presentes na TAB. 28, percebemos que há uma grande variação no percentual de cancelamento. Há palavras que apresentam um alto percentual de cancelamento, como *sempre* (com 50%) e *eles* (com 42,9%). Contudo, há palavras com até 0% de cancelamento, como *grande* e *maior*. Essa diferença é encontrada, inclusive, em palavras relacionadas a um mesmo fenômeno, como as citadas *sempre* (com 50%) e *grande* (com 0%) que se relacionam ao Fenômeno 3 (cc). Possivelmente, a razão dessa diferença entre *sempre* e *grande* se deve à questão da tonicidade. Isso porque, em *sempre*, o som analisado se encontra em sílaba postônica (*sem-pre*), que é o contexto

que mais favorece o Fenômeno 3 (cc) (CRISTÓFARO SILVA, 2003). Ao contrário, *grande* apresenta o som em análise em sílaba tônica (*gran-de*), contexto este que tende a desfavorecer tal fenômeno (CRISTÓFARO SILVA, 2003). Novamente, vemos aqui a importância de se avaliar o fator estrutural, o qual pode atuar como um estabilizador da mudança sonora difundida lexicalmente (OLIVEIRA, M., 1992).

A fim de visualizar como essas 16 palavras se agrupam em termos de seus índices de cancelamento, apresentamos uma análise de classificação em árvore (método CHAID) na FIG. 38 que demonstra isso:

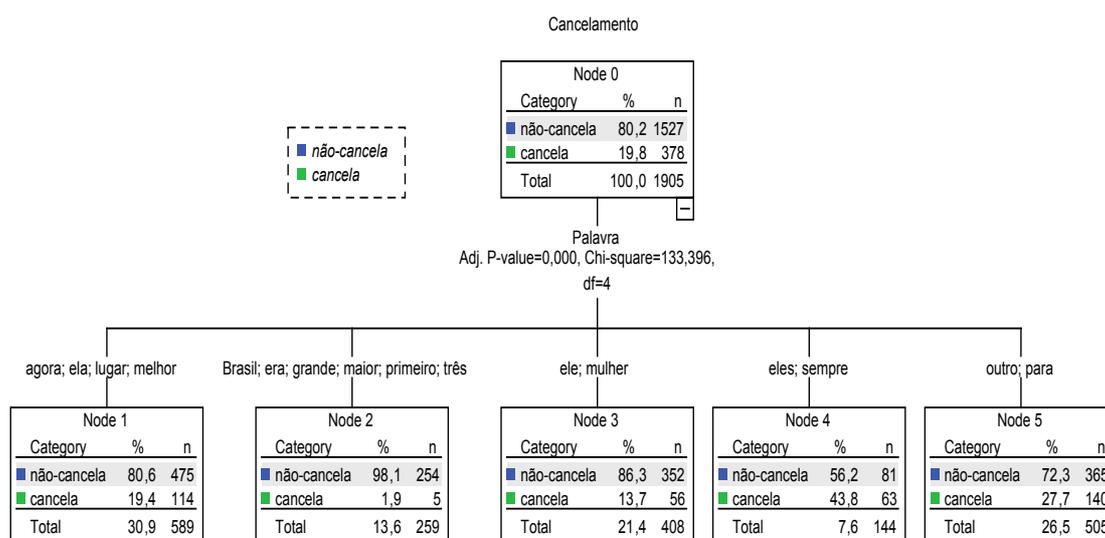


FIGURA 38 – Classificação em árvore (método CHAID) para as 16 palavras consideradas no estudo longitudinal

A FIG. 38 nos mostra, primeiramente, que as 16 palavras avaliadas no estudo longitudinal foram divididas em 5 nós/ grupos de acordo com a homogeneidade que há entre elas em termos de índice de cancelamento. Assim, por essa figura, percebemos

que as palavras que apresentam maior percentual médio de cancelamento, com uma média de 43,8%, são *eles* e *sempre* (4º nó).

Em segundo lugar, vêm as palavras *outro* e *para* (5º nó), com uma média de 27,7% de cancelamento. Em terceiro, vêm as palavras *agora*, *ela*, *lugar* e *melhor* (1º nó), com 19,4%. Em quarto, vêm as palavras *ele* e *mulher* (3º nó) com 13,7%. Em quinto e último lugar, vêm as palavras *Brasil*, *era*, *grande*, *maior*, *primeiro*, *três* (2º nó), com o mais baixo percentual médio de 1,9% de cancelamento.

É preciso ressaltar que, das 6 palavras deste último grupo, 4 delas – ou seja, a maioria – se referem ao Fenômeno 3 (cc): *Brasil*, *grande*, *primeiro*, *três*. Por outro lado, das 4 palavras dos dois grupos que apresentam maior percentual médio de cancelamento (*eles*, *sempre* no 4º nó e *outro*, *para* no 5º nó), 2 se referem a esse fenômeno: *sempre* (4º nó) e *outro* (5º nó). Com isso, vemos que há variabilidade no comportamento das palavras em análise mesmo quando elas são observadas em relação a um único fenômeno. Esse resultado, mais uma vez, aponta para o fato de a mudança sonora ser lexicalmente gradual (CHEN; WANG, 1975).

Vale ainda destacar que a explicação para tal variabilidade pode estar ligada a fatores estruturais. Quanto a essas 4 palavras do Fenômeno 3 (cc) que apresentam baixo percentual médio de cancelamento: a) em duas palavras, *grande* e *três*, o som a ser cancelado incide sobre a sílaba tônica, que tende a desfavorecer fenômenos de cancelamento sonoro (CRISTÓFARO SILVA, 2003; FONTES MARTINS, 2001); b) uma palavra, *Brasil*, é nome próprio, que, como vimos com Oliveira, M. (1991), também tende a desfavorecer a ocorrência de fenômenos de mudança sonora; e c) a última palavra, *primeiro*, apresenta o som a ser cancelado em início de palavra, em

sílaba pretônica, a qual favorece menos tal fenômeno do que a sílaba postônica (CRISTÓFARO SILVA, 2003). Já quanto às palavras que mais cancelaram, *sempre* e *outro*, vemos que, em ambas, o som a ser cancelado encontra-se em sílaba postônica, que é o contexto que mais favorece fenômenos de cancelamento sonoro (CRISTÓFARO SILVA, 2003; FONTES MARTINS, 2001).

Por toda essa análise relacionada à *Palavra*, vemos que são compatíveis as idéias de que a mudança sonora é difundida de modo lexicalmente gradual (CHEN; WANG, 1975) e que o fator estrutural tem sua influência, ao atuar como um estabilizador de tal mudança (OLIVEIRA, M., 1992). Passemos, a seguir, a fazer a análise das 16 palavras do estudo longitudinal para cada um dos indivíduos avaliados neste estudo, L2R e N3Z, em separado. Para tanto, utilizaremos a análise de classificação em árvore (método CHAID), a fim de verificar como tais palavras se distribuem com relação aos dois indivíduos<sup>58</sup>. Vejamos, assim, as FIG. 39 e FIG. 40, as quais se relacionam aos Indivíduos L2R e N3Z, respectivamente:

---

<sup>58</sup> Também aqui gostaríamos de ter feito análises de regressão logística binária, contudo, isso não foi possível, pelo fato de os dados se encontrarem mal distribuídos. Para consulta, dispomos, no APÊNDICE G, tabelas demonstrando os percentuais de cancelamento de cada palavra para os dois indivíduos do estudo longitudinal. Novamente aqui é bom ressaltar que, pelas análises do APÊNDICE G, vemos que um mesmo indivíduo pode ter produções diferentes para uma única palavra. Mesmo o Indivíduo não-líder – N3Z, que é muito conservador, apresenta, para as 9 ocorrências da palavra *sempre*, 5 ocorrências canceladas e 4 não-canceladas. Esse fato demonstra variabilidade intra-individual, o que se contrapõe ao estudo de Oliveira, M. (2006).

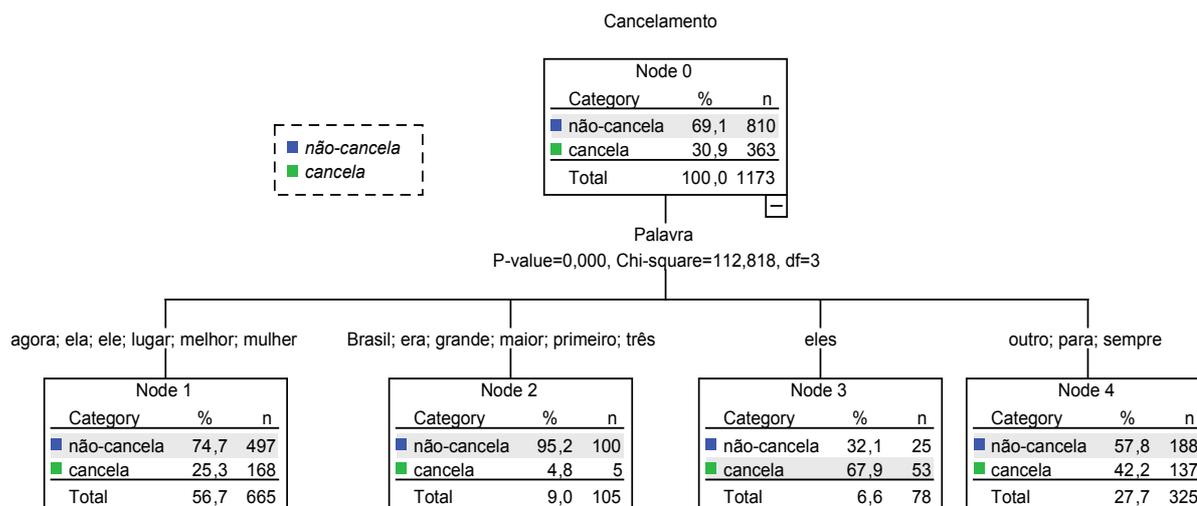


FIGURA 39 – Indivíduo – L2R e as 16 palavras consideradas no estudo longitudinal

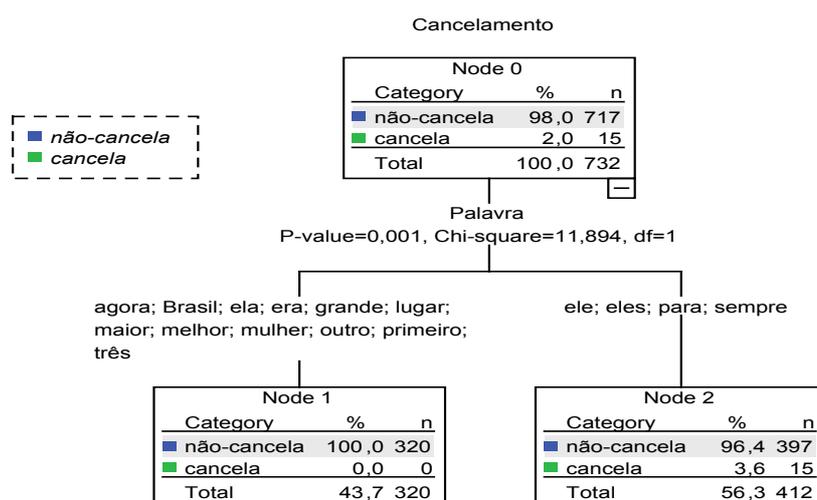


FIGURA 40 – Indivíduo – N3Z e as 16 palavras consideradas no estudo longitudinal

Comparando as FIG. 39 e FIG. 40, logo de início percebemos que o agrupamento das palavras é diferente para cada um dos dois indivíduos: enquanto L2R as agrupa em 4 nós (que apresentam uma média de 4,8% a 67,9% de cancelamento), N3Z as agrupa em 2 nós apenas (um com uma média de 3,6% de cancelamento e o outro com 0%). Como os percentuais de cancelamento médio dos dois indivíduos são muito discrepantes (por exemplo, enquanto L2R tem como máximo 67% de cancelamento, N3Z tem 3,6%

apenas), fica difícil comparar L2R e N3Z quanto à variável *Palavra*. Contudo, fato é que essa grande diferença entre L2R e N3Z quanto às 16 palavras nos permite concluir que eles apresentam, no uso, léxicos diferentes (BYBEE, 2001; REINECKE, 2006), que eles avaliam de forma muito diversa tais palavras.

Apesar dessa discrepância entre L2R e N3Z, guardadas as devidas proporções, há uma semelhança entre eles. Três das quatro únicas palavras do Indivíduo – N3Z que apresentam cancelamento – *eles, para, sempre* (com 3,6% de cancelamento, Cf. 2º nó da FIG. 40) – estão compreendidas nas palavras do Indivíduo – L2R que apresentam maior índice de cancelamento médio (Cf. *eles* com 67,9% de cancelamento no 3º e *para, sempre* com 42,2% no 4º nó da FIG. 39). Com isso, vemos que possivelmente, para esses dois indivíduos, as palavras *eles, para* (Fenômeno 2 (v-v)) e *sempre* (Fenômeno 3 (cc)) estão lexicalmente mais difundidas do que as outras relacionadas aos seus respectivos fenômenos (CHEN; WANG, 1975).

#### 6.4 Sumário

Neste capítulo, primeiramente, na seção 6.2, abordamos uma análise que fizemos, a qual teve um caráter longitudinal, visto que, em tal análise, avaliamos se houve mudança estatisticamente significativa, para os 12 indivíduos em estudo, do corpus de origem de cada um deles (corpora 1, 2 e 3) para o corpus atual. Como resultado, vimos que houve variabilidade intra-individual significativa (SANGSTER, 2002) em 7 desses 12 indivíduos, com relação ao tempo que se passou das coletas das pesquisas anteriores para a coleta da pesquisa atual (PAIVA; DUARTE, 2003).

Na seção 6.3, apresentamos e discutimos os resultados referentes aos dados do estudo longitudinal realizado com os indivíduos líder L2R e não-líder N3Z. Nessa análise, foram avaliadas, além da variável dependente *Cancelamento*, as variáveis independentes: *Indivíduo*, *Coleta* e *Palavra*.

Ao abordamos a variável *Indivíduo*, vimos que, quanto aos dois indivíduos avaliados no estudo longitudinal, a informante L2R, que é líder, se comporta como líder realmente, assim como N3Z, que é não-líder, se comporta como não-líder de modo geral, nas 6 coletas realizadas no estudo longitudinal (LABOV, 2001, v. 1). Contudo, ponderamos que, considerando os efeitos de protótipo de Pierrehumbert (2001b), enquanto a líder – L2R não pode ser considerada como o protótipo da sua categoria, “líder” – por não apresentar índices tão altos de realização dos fenômenos em análise –, ao contrário, o não-líder – N3Z pode ser considerado o protótipo de sua categoria, “não-líder”, por apresentar os menores índices de realização de tais fenômenos tanto no nosso corpus do estudo transversal, quanto do longitudinal (PIERREHUMBERT, 2001b). Destacamos que esse fato, inclusive, aponta para a distribuição probabilística das categorias “líder” e “não-líder” em um contínuo gradual (PIERREHUMBERT, 2001b; BOD; HAY; JANNEDY, 2003).

Ao analisar a variável *Coleta*, vimos que o fator tempo, implícito nas diferentes coletas, não foi significativo em nossa análise, já que não houve diferença estatisticamente relevante para L2R e N3Z, quanto às 6 coletas (PAIVA; DUARTE, 2003). Isso pode ter ocorrido pelo fato de o tempo médio entre as coletas (um mês) e o tempo total de coleta (6 meses) terem sido pequenos para se observarem mudanças significativas no comportamento dos indivíduos L2R e N3Z quanto aos fenômenos em estudo. Também,

isso pode ter ocorrido pelo fato de esses indivíduos não terem passado por mudanças, por exemplo, sócio-emotivas, significativas em suas vidas (como, casar-se ou ter filho(s), ou mudar de casa, de emprego e/ ou de escolaridade, etc.) nesse intervalo de tempo (SANGSTER, 2002; PAIVA; DUARTE, 2003). Como apontamos, esse resultado é interessante, por demonstrar que, mesmo com variabilidade, há estabilidade na linguagem adulta, sendo a estabilidade crucial para a manutenção da linguagem de geração para geração.

A respeito da variável *Palavra*, destacamos que fizemos um recorte dessa variável, o que reduziu as 1.646 categorias/ tipos de palavra para apenas 16 categorias/ tipos de palavra (BYBEE, 2001). Ao analisar essas 16 palavras, vimos que elas também apresentavam percentuais de cancelamento diferentes, mesmo palavras relativas a um único fenômeno. Por fim, ao fazer uma comparação entre os indivíduos L2R e N3Z quanto a essas 16 palavras, vimos que eles as avaliam de forma muito diversa, o que aponta para o fato de que, no uso, indivíduos apresentam léxicos e gramáticas diferentes (BYBEE, 2001; REINECKE, 2006). Concluimos que os resultados relativos à palavra apontam para a fundamental importância desta (CHEN; WANG, 1975; BYBEE, 2001) em fenômenos de mudança sonora e para a dinamicidade e a plasticidade do léxico (BYBEE, 2001; PIERREHUMBERT, 2001b). No próximo capítulo, passamos às considerações finais deste trabalho.

## CAPÍTULO 7

### CONSIDERAÇÕES FINAIS

Por meio da análise do comportamento lingüístico de indivíduos específicos em relação a fenômenos fonológicos diferentes, este trabalho buscou contribuir para a discussão sobre a organização do componente fonológico.

Assim, avaliando três fenômenos fonológicos, esta pesquisa teve como principal objetivo analisar a variabilidade individual, para uma maior compreensão da natureza das representações lingüísticas – enfocando as representações fonológicas – organizadas probabilisticamente. Sugerimos que a variabilidade atestada na produção de fala dos indivíduos (variações inter e intra-individuais) está contida nas representações lingüísticas deles. Desse modo, adotamos como quadro teórico as seguintes teorias multirrepresentacionais: Fonologia de Uso (BYBEE, 2001), Teoria de Exemplos (PIERREHUMBERT, 2001b, 2003a) e Lingüística Probabilística (BOD; HAY; JANNEDY, 2003).

Discutimos os estudos de Huback (2003), Oliveira, M. (1992) e Oliveira Guimarães (2004), os quais atestam variações interindividuais em relação a fenômenos fonológicos e por isso sugerem que se considere o comportamento do indivíduo na análise de tais fenômenos. Abordamos os trabalhos de Paiva e Duarte (2003) e Shen (1990), os quais apontam para a difusão gradual da mudança sonora na comunidade, que atinge falante a falante. Destacamos que Paiva e Duarte (2003), ao reconhecerem irregularidades no comportamento do indivíduo, enfatizam as variações intra-individuais.

Tratamos de três abordagens para fenômenos de mudança sonora: 1) comunidade lingüística (LABOV, 1966); 2) redes sociais (MILROY, 1987); 3) comportamento do indivíduo (SANGSTER, 2002). Após avaliar os conceitos de comunidade lingüística e de redes sociais, já dentro da terceira abordagem do comportamento do indivíduo, discutimos o conceito de líder da mudança lingüística, proposto por Labov (2001, v. 2), conceito este que serviu como critério para a seleção dos participantes desta pesquisa. Ainda nessa terceira abordagem, tratamos do trabalho de Sangster (2002), a qual encontrou variações inter e intra-individuais na análise de diferentes estilos e fenômenos fonológicos. Com base no estudo dessa autora, propusemos, dentro dos objetivos desta tese, investigar os indivíduos em análise quanto a três estilos e três fenômenos fonológicos.

Ao esclarecer os procedimentos metodológicos adotados nesta pesquisa, destacamos que não existe uma metodologia específica para o tratamento da variabilidade individual. A metodologia aqui proposta visou ser a mais adequada possível a esse tratamento, sendo uma contribuição desta pesquisa para trabalhos futuros que abordem o comportamento do indivíduo. Propusemos dois tipos de estudo para abordar o tema desta pesquisa: um transversal com 12 indivíduos (6 líderes e 6 não-líderes) e um longitudinal de 6 coletas mensais com 2 desses 12 indivíduos (1 líder e 1 não-líder).

Retomemos, neste momento, os objetivos específicos propostos pelo presente trabalho (Cf. capítulo 1), os quais se relacionam a esses dois estudos:

1. caracterizar o comportamento do indivíduo quanto a diferentes fenômenos fonológicos do português brasileiro, mais especificamente, quanto aos três fenômenos;
2. avaliar a variabilidade individual com relação a itens específicos, ou seja, quanto ao fator palavra;
3. caracterizar o comportamento individual com relação a diferentes estilos;
4. avaliar a proposta de Labov (2001, v. 2) quanto à questão da liderança em casos de mudança sonora;
5. analisar, com relação a fenômenos de mudança sonora, o comportamento de indivíduos adultos em estudo longitudinal (PAIVA; DUARTE, 2003).

Tais objetivos se relacionam às variáveis analisadas nos estudos transversal e longitudinal. Tratando primeiramente de cada variável do estudo transversal, na análise dos dados, avaliamos estas 5 variáveis: 1) *Fenômeno*; 2) *Palavra*; 3) *Estilo*; 4) *Liderança*; e 5) *Indivíduo*. Aprofundando tal análise, avaliamos as variáveis *Liderança* e *Indivíduo*, relacionando-as com *Fenômeno*, *Palavra* e *Estilo*.

Ao avaliar a variável *Fenômeno*, que se relaciona ao objetivo (1) listado acima, sugerimos que o Fenômeno 1 (r #) parece estar lexicalmente mais difundido (CHEN; WANG, 1975; BYBEE, 2001) do que os Fenômenos 2 (v-v) e 3 (cc). Na análise da variável *Palavra*, que se relaciona ao objetivo (2), vimos que os resultados apontam para a dinamicidade e a plasticidade do léxico (OLIVEIRA, M., 1995; BYBEE, 2001; PIERREHUMBERT, 2001b), já que palavras relacionadas a um mesmo fenômeno demonstraram padrões de difusão lexical da mudança sonora (BYBEE, 2001; CHEN; WANG, 1975).

Na análise da variável *Estilo*, que se relaciona ao objetivo (3), constatamos uma gradiência (PIERREHUMBERT, 2002) entre os três diferentes estilos analisados, o que aponta para o caráter probabilístico da linguagem (BOD; HAY; JANNEDY, 2003). Destacamos o fato de que tipos de estilo podem ser estatisticamente diferentes quanto a um fenômeno, mas não quanto a um outro, e que a *fala espontânea* atua como “motor” da variabilidade.

Ao avaliar a variável *Liderança*, que se relaciona ao objetivo (4), com relação a todos os 3 fenômenos ao mesmo tempo, verificamos que existe um comportamento de tendência, segundo o qual, em conjunto, os líderes se comportam como líderes, assim como também em conjunto, os não-líderes se comportam como não-líderes (LABOV, 2001, v. 2). No entanto, vimos que, ao avaliar os três fenômenos separadamente, na análise da relação entre *Liderança* e *Fenômeno*, líderes e não-líderes apresentaram padrões diferentes para tais fenômenos (BYBEE, 2001; SANGSTER, 2002): enquanto os líderes mantiveram índices de cancelamento parecidos em cada um dos fenômenos, os não-líderes apresentaram índices estatisticamente diferentes. Na análise da relação entre *Liderança* e *Palavra*, observamos que, no uso, líderes apresentam maior variabilidade e multiplicidade de exemplares de uma mesma palavra (BYBEE, 2001; PIERREHUMBERT, 2001b) do que não-líderes. Já na análise da relação da *Liderança* com *Estilo*, notamos que, enquanto para os líderes há uma distinção bem nítida entre estilo informal (*fala espontânea*) e estilo formal (*leitura e nomeação de figura*), para os não-líderes não há, visto que, na análise estatística, um estilo informal (*fala espontânea*) foi agrupado com um estilo formal (*nomeação de figura*) (LABOV, 1972; ABAURRE, 1981; SANGSTER, 2002). Com esse resultado, sugerimos que, para os não-líderes, a *fala espontânea* parece ser um estilo mais formal do que informal (LABOV, 1972; MARCHUSCHI, 2003).

Na análise da última variável do estudo transversal, *Indivíduo*, a qual se relaciona ao objetivo geral desta pesquisa, enfatizamos a importância de se considerar o indivíduo como unidade de análise (OLIVEIRA, M., 1992; SANGESTER, 2002), já que encontramos, dentro das categorias *líder* e *não-líder*, variabilidade individual significativa segundo a análise estatística realizada. Ao avaliar a relação entre *Indivíduo* e *Fenômeno*, vimos que os indivíduos lidam com diferentes fenômenos de forma probabilística (BOD; HAY; JANNEDY, 2003). Ressaltamos a importância de se considerar fenômenos diferentes para os indivíduos (SANGSTER, 2002), a fim de se observarem variações inter e intra-individuais. Esta é uma sugestão metodológica deste trabalho para estudos futuros que se ocupem em analisar tais variações. Ao avaliar a relação *Indivíduo* e *Palavra*, observamos que a organização do componente fonológico varia individualmente em função da *Palavra* e que isso faz com que indivíduos diferentes apresentem léxicos e gramáticas diferentes no uso (HOPPER, 1987; BYBEE, 2001; REINECKE, 2006). Na análise da relação *Indivíduo* e *Estilo*, destacamos que a organização do componente fonológico, mais uma vez, pode variar individualmente, de acordo com o estilo (SANGSTER, 2002).

Passando ao estudo longitudinal, primeiramente abordamos uma análise, que se relaciona ao objetivo (5), em que avaliamos se houve mudança estatisticamente significativa, para os 12 indivíduos em estudo, do corpus de origem de cada um deles (corpora 1, 2 e 3) para o corpus atual. Como resultado, vimos que houve variabilidade intra-individual significativa (SANGSTER, 2002) em 7 desses 12 indivíduos, com relação ao tempo que se passou das coletas das pesquisas anteriores para a coleta da pesquisa atual (PAIVA; DUARTE, 2003). Em seguida, apresentamos e discutimos os resultados referentes aos dados do estudo longitudinal realizado com os indivíduos líder

L2R e não-líder N3Z. Nessa análise, foram avaliadas, além da variável dependente *Cancelamento*, as variáveis independentes: *Indivíduo*, *Coleta* e *Palavra*.

Ao abordar a variável *Indivíduo*, vimos que a informante L2R, que é líder, se comporta como líder, assim como N3Z, que é não-líder, se comporta como não-líder, nas 6 coletas realizadas no estudo longitudinal (LABOV, 2001, v. 2). Ao analisar a variável *Coleta*, vimos que o fator tempo, implícito nas diferentes coletas, não foi significativo, já que não houve diferença estatisticamente relevante para L2R e N3Z, quanto às 6 coletas realizadas (PAIVA; DUARTE, 2003). Sugerimos que isso ocorreu pelo fato de o tempo médio entre as coletas (um mês) e o tempo total de coleta (6 meses) terem sido pequenos para se observarem mudanças significativas no comportamento dos Indivíduos L2R e N3Z quanto aos fenômenos em estudo. Também, isso pode ter ocorrido pelo fato de esses indivíduos não terem passado por mudanças, por exemplo, sócio-emotivas, significativas em suas vidas (como, casar-se ou ter filho(s), ou mudar de casa, de emprego e/ ou de escolaridade, etc.) nesse intervalo de tempo (SANGSTER, 2002; PAIVA; DUARTE, 2003). Como apontamos, esse resultado é interessante, por demonstrar que há uma certa estabilidade na linguagem adulta.

A respeito da variável *Palavra*, vimos que, assim como no estudo transversal, as palavras analisadas, mesmo quando se considera um único fenômeno, apresentaram percentuais de cancelamento diferentes, demonstrando também padrões de difusão lexical (CHEN; WANG, 1975; BYBEE, 2001). Por fim, ao fazer uma comparação entre os indivíduos L2R e N3Z quanto a essas 16 palavras, vimos que eles as avaliam de forma muito diversa, o que aponta para o fato de que, no uso, indivíduos têm léxicos e gramáticas diferentes (BYBEE, 2001; REINECKE, 2006).

Os resultados desta pesquisa evidenciam o cumprimento dos objetivos aqui propostos e apontam para o fato de que o estudo do comportamento do indivíduo é de grande relevância para uma compreensão mais ampla da organização do componente lingüístico e do correlato social da linguagem (OLIVEIRA, M., 1992; SANGSTER, 2002; NARO, 2003; PAIVA; DUARTE, 2003). Pesquisas futuras poderiam investigar, mais amplamente, as causas das variações inter e intra-individuais. A variabilidade individual poderia ser mais explorada em outras áreas, como a da aquisição da linguagem (fala infantil). Ainda, trabalhos posteriores poderiam pesquisar variações no comportamento do indivíduo quanto:

- à aquisição e ao desenvolvimento de segunda língua (JENKINS, 2000);
- à aquisição e ao desenvolvimento da escrita (SOARES, 2003);
- à gradiência fonética dos dados (PIERREHUMBERT, 2002), explorando análises acústicas;
- a padrões estruturais de cancelamento/ lenição em rede (BYBEE, 2001);
- a fatores relativos à questão da liderança, como a investigação sobre a ambição e os anseios de vida dos indivíduos (LABOV, 2001, v. 2; SANGSTER, 2002);
- à relação entre taxa de elocução e redução segmental (BARBOSA, 2000) para líderes e não-líderes (LABOV, 2001, v. 2);
- à relação entre redução da magnitude dos gestos articulatórios (ALBANO, 2001) e redução segmental;
- à audiência, tendo em vista que Sangster (2002) encontrou variações inter e intra-individuais relacionadas à audiência em seus dados do inglês. Neste estudo, não pudemos avaliar esse fator, pois, na coleta de nossos dados, a interação dos participantes foi somente com a presente pesquisadora, sendo que,

para avaliar a audiência, seriam necessárias interações com mais de um pesquisador para cada um dos participantes;

- a mudanças sócio-emotivas na vida dos indivíduos – mudança de escolaridade, de emprego, de residência; maternidade/ paternidade; viagem para o exterior; doença, etc. (SANGSTER, 2002; PAIVA; DUARTE, 2003).

Por fim, resta destacar que, nesta tese, não pretendemos esgotar todas as possibilidades de análise para o comportamento do indivíduo em relação ao componente fonológico. Trabalhos como os sugeridos na área da fonologia e, mesmo, em outras áreas da Lingüística seriam de grande valia para uma maior compreensão da variabilidade individual e da organização e uso da gramática.

**BIBLIOGRAFIA**

ABAURRE, M. B. Processos fonológicos segmentais como índices de padrões prosódicos diversos nos estilos formal e casual do português do Brasil. *Caderno de Estudos Lingüísticos*, 2, p. 23-34, 1981.

ALBANO, E. C. *O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado de Letras, 2001. 272 p.

AMARAL, Marisa P. *As proparoxítonas: teoria e variação*. 2001. Tese (Doutorado em Lingüística) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto alegre, 1999.

ARAÚJO, Gabriel Antunes (Org.). *O acento em português: abordagens fonológicas*. São Paulo: Parábola, 2007.

ASH, Sharon. Social Class. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002. p. 402-422.

BAGNO, Marcos. *Preconceito lingüístico: o que é, como se faz*. São Paulo: Edições Loyola, 2002.

BAILEY, Guy. Real and Apparent Time. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002. p. 312-332.

BALEN, S. A. *Análise por traços distintivos do sistema fonológico de crianças com alterações na fala*. 1995. Monografia (Especialização) – Universidade Federal de Santa Maria, Santa Maria, RS, 1995.

BALEN, S. A. *Processamento auditivo central: aspectos temporais da audição e percepção acústica da fala*. 1997. Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1997.

BARBOSA, Plínio Almeida. Syllable-timing in Brazilian Portuguese: uma crítica a Roy Major. *D.E.L.T.A.*, v. 16, v. 2, p. 369-402, 2000.

BERBER SARDINHA, Tony. Lingüística de corpus: histórico e problemática. *D.E.L.T.A.*, v. 16, n. 2, p. 323-367, 2000.

BICKERTON, D. *Dynamics of a Creole System*. Cambridge: Cambridge University Press, 1975.

BISOL, Leda. O Troqueu Silábico no Sistema Fonológico. *D.E.L.T.A.*, v. 16, v. 2, 2000. Disponível em: < [http://www.scielo.br/scielo.php?pid=S0102-44502000000200007&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0102-44502000000200007&script=sci_arttext)>. Acesso em: 02 mar 2007.

BLOCH, Bernard. A set of postulates for phonemic analysis. *Language*, 24, p. 3-46, 1948.

BOD, J.; HAY, J.; JANNEDY, S (Eds.). *Probabilistic Linguistics*. Cambridge: MIT Press, 2003.

BOD, Rens. Introduction to Probability theory in linguistics. LSA – Workshop *Probability Theory in Linguistics*. Washington D.C. Linguistic Society of America, 2001.

BONILHA, Giovana F. G. *Aquisição fonológica do português brasileiro: uma abordagem conexionista da teoria da otimidade*. 2004. Tese (Doutorado em Lingüística) – Faculdade de Letras, PUCRS, Porto Alegre, 2004.

BROWMAN, C. P.; GOLDSTEIN, L. Articulatory phonology: an overview. *Phonetica*, v. 49, p. 155-180, 1992.

BUCCI, Wilma. Linking words and things: basic processes and individual variation. *Cognition*, n. 17, p. 137-153, 1984.

BYBEE, Joan. Regular Morphology and the Lexicon. *Language and Cognitive Processes*. U. K.: Erlbaum, 1995.

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

BYBEE, Joan. Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change. *Language Variation and Change*, v. 14, p. 261-290, 2002.

BYBEE, Joan. The impact of us on representations: Grammar is usage and usage is grammar. LSA Presidential Address, 2005.

BYBEE, Joan. From usage to grammar: the mind's response to repetition. *Language* 82, 4, p. 529-551, 2006.

CARDOSO, Caroline Rodrigues. *Variação da concordância verbal no indivíduo: um confronto entre o lingüístico e o estilístico*. 2005. Dissertação (Mestrado em Lingüística) – Instituto de Letras, UNB, Brasília, 2005.

CHAMBERS, J. K. *Sociolinguistic Theory*. Cambridge: Blackwell, 1995.

CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002.

CHEN, Matthew Y.; WANG, William S-Y. Sound Change: actuation and implementation. *Language*, v. 51, n. 2, p. 255-281, 1975.

CHOMSKY, Noam. *Aspects of the theory syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, Noam; HALLE, Morris. *The Sound Pattern of English*. New York: Harper & Row, 1968.

CHOMSKY, Noam. Conhecimento da História e construção Teórica na Lingüística Moderna. *DELTA*, v. 13, número especial, p. 133-155, 1997. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44501997000300005&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44501997000300005&lng=pt&nrm=iso). Acesso em: 24 maio 2005.

COLLISCHONN, G. *Um Estudo do Acento Secundário em Português*. Dissertação (Mestrado) – UFRGS, Porto Alegre, 1993.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). Diretório dos Grupos de Pesquisa no Brasil, versão 5.0 (Censo 2002). Disponível em: <http://lattes.cnpq.br/censo2002/>. Acesso em: 20 ago. 2004.

CORPUS DE EXTRATOS DE TEXTOS ELETRÔNICOS (CETEM). Disponível em: <http://www.linguateca.pt/CETEMPublico/>. Acesso em: fev. e mar. 2004.

CRISTÓFARO SILVA, Thaís. Sobre a queda dos encontros consonantais no português brasileiro. *Estudos lingüísticos XXIX*. v. 29, 2000 a.

CRISTÓFARO SILVA, Thaís. Branching Onsets in Brazilian Portuguese. 30<sup>th</sup> Linguistic Symposium in Romance Languages. University of Florida. Gainesville, 2000 b.

CRISTÓFARO SILVA, Thaís. Descartando Fonemas: a representação lexical na 'Fonologia de Uso'. In: DA HORA, Dermeval; COLLISCHONN, Gisella. (Org.). *Teoria Lingüística: Fonologia e Outros Temas*. João Pessoa: Universidade Federal da Paraíba, 2003, p. 200-231.

CRISTÓFARO SILVA, Thaís. *Difusão Lexical: estudo de casos do Português Brasileiro*. Manuscrito. Belo Horizonte: Faculdade de Letras da UFMG, 2003. Relatório de Pesquisa CNPq número: 202422-86-3.

CRISTÓFARO SILVA, Thaís. Cluster reduction in Brazilian Portuguese. Handout apresentado no *LAGB Spring Meeting*, 2003b, University of Sheffield. Disponível em: <<http://www.lettras.ufmg.br/cristofaro/publicacoes/2003lagb.pdf>>. Acesso em: 10 ago. 2004.

CRISTÓFARO SILVA, Thaís; ABREU GOMES, Christina. Representações múltiplas e organização do componente fonológico. Artigo submetido ao *Fórum Lingüístico* da UFSC, 2004.

CRYSTAL, David. *Dicionário de Lingüística e Fonética*. Rio de Janeiro: Jorge Zahar Editor, 1988.

DEMUTH, K. Multiple optimal outputs in acquisition. University of Maryland. Working papers in linguistics, n. 5, p. 53-71, 1997.

DOCHERTY, Gerard J. Speaker, Community, Identity: Empirical and Theoretical Perspectives on Sociophonetic Variation. *ICPHS3*, University of Newcastle upon Tyne, UK, 2003.

DOCHERTY, Gerard J. *et al.* Descriptive adequacy in phonology: a variationist perspective. *Linguistics*, v. 33, p. 275-310, 1997.

ECKERT, Penélope. Constructing meaning in sociolinguistic variation. Paper presented at the *Annual Meeting of the American Anthropological Association*. New Orleans, nov. 2002.

ECKERT, Penélope; RICKFORD, John R. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 2002.

ELLIS, Nick C. Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *SSLA*, v. 24, Cambridge University Press, p. 143-188, 2002.

EVANS, Betsy. The role of social network in the acquisition of local dialect norms by Appalachian migrants in Ypsilanti, Michigan. *Language Variation and Change*, v. 16, Cambridge University Press, p. 153-167, 2004.

FARACO, C. A. *Lingüística histórica: uma introdução ao estudo da história das línguas*. São Paulo: Ática, 1998.

FIDELHOLTZ, J. Word frequency and vowel reduction in English. *Papers from the eleventh Regional Meeting Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1975. p. 200-213.

FILLMORE, Charles J.; KEMPLER, Daniel; WANG, William S-Y. *Individual differences in language ability and language behavior*. New York: Academic Press, 1979.

FONTES MARTINS, Raquel M. *O cancelamento das líquidas /l/ e /r/ intervocálicas no português de Belo Horizonte*. 2001. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2001.

FONTES MARTINS; OLIVEIRA GUIMARÃES (Em andamento). A organização do componente fonológico e o comportamento do indivíduo: uma proposta de estudo longitudinal. In: VII ENCONTRO NACIONAL SOBRE AQUISIÇÃO DA LINGUAGEM – ENAL 2006. Porto Alegre. Em andamento.

FRANÇA, Júnia Lessa *et al.* *Manual para normalização de publicações técnico-científicas*. 7 ed. Belo Horizonte: UFMG, 2004.

FREITAS, Edir Ramos de. *Aprendizagem da estrutura silábica CCV: oralidade e escrita*. 2001. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2001.

FRIDLAND, Valerie; BARTLETT, Kathry; KREUZ, Roger. Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN. *Language Variation and Change*, v. 16, Cambridge University Press, p. 1-16, 2004.

GOLDINGER, S. D. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 22, p. 1166-1183, 1996.

GOLDSTONE, Robert L. The role of similarity in categorization: providing a groundwork. *Cognition*, 57, p. 125-157, 1994.

GORDON, Barry; CARAMAZZA, Alfonso. Lexical access and frequency sensitivity: frequency saturation and open/closed class equivalence. *Cognition*, v. 21, p. 95-115, 1985.

GUMPERZ, J. J. The sociolinguistic significance of conversational codeswitching. *Working Paper 46*, Language Behaviour Research Laboratory, Berkeley, 1976b.

HARRIS, John. Towards a lexical analysis of sound change in progress. *Linguistics* 25. London: Department of Phonetics and Linguistics – University College London, p. 35-56, 1989.

HARRIS, John. *English sound structure*. Cambridge: Blackwell, 1994.

HAY, Jennifer. Lexical frequency in morphology: is everything relative? *Linguistics* v. 39, n. 6, p. 1041–1070, 2001.

HAZAN, V.; BARRETT, S. The development of phonemic categorization. *Journal of phonetics*, v. 28, p. 377-396, 2000.

HAZEN, Kirk. The Family. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002. p. 500-525.

HERNANDORENA, C. L. M. *Aquisição da fonologia do Português: estabelecimentos dos padrões com base em traços distintivos*. 1990. Tese (Doutorado) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1990.

HOCKETT, C. F. *A Course in Modern Linguistics*. New York: Macmillan, 1958.

HOLANDA, Aurélio Buarque de. *Novo Dicionário Aurélio da Língua Portuguesa*. São Paulo. Disponível em: <<http://www2.uol.com.br/aurelio/index.html>>. Acesso em: 27 fev. 2004.

HOLANDA, Aurélio Buarque de. *Novo Dicionário Eletrônico Aurélio: versão 5.11a*. CDROM. 3 ed. São Paulo: Positivo, 2004.

HOPPER, Paul J. Emergent grammar. *Berkeley Linguistics Society*, vol. 13, 1987, p. 139-157.

HUBACK, Ana Paula da Silva. *Cancelamento do (R) final em nominais na cidade de Belo Horizonte: uma abordagem difusionista*. 2003. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2003.

JENKINS, J. *The phonology of english as an Internacional language*. Oxford: OUP, 2000.

JOHANSEN, Mark K.; PALMERI, Thomas J. Are there representational shifts during category learning?. *Cognitive Psychology*, v. 45, p. 482-553, 2002.

JOHNSON, K. Speech perception without speaker normalization. In: JOHNSON, K.; MULLENNIX (Eds.). *Talker variability in Speech Processing*. San Diego: Academic Press, 1997.

JOHNSON, K.; MULLENNIX (Eds.). *Talker variability in Speech Processing*. San Diego: Academic Press, 1997.

JOHNSTONE, Barbara. *The linguistic individual: Self-Expression in languages and linguistics*. Oxford: Oxford University Press, 1997.

JOHNSTONE, Barbara. *Qualitative Methods in Sociolinguistics*. Oxford: Oxford University Press, 2000.

JOHNSTONE, Barbara; BEAN, Judith Mattson. Self-expression and linguistic variation. *Language in Society*, v. 26, p. 221-246, 1997.

JONES, Mari C.; ESCH, Edith. *Language Change: the interplay of internal, external and extra-linguistic factors*. Berlin, New York: Mouton de Gruyter, 2002.

KINNEAR, Paul R.; COLIN, D. Gray. *SPSS for Windows Made Simple: Release 10*. East Sussex: Psychology Press. 2001.

LABOV, W. *The social stratification of English in New York city*. Arlington: Center for Applied Linguistics, 1966.

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

LABOV, W. *The design of sociolinguistic research project*. [This article is the chapter II of the report of sociolinguistics workshop held by the Central Institute of Indian Languages in Mysore, India, May-June, 1972], 1974.

LABOV, W. *Language in the inner city*. Philadelphia: Pennsylvania University Press, 1975.

LABOV, W. Exact description of the speech community: Short *a* in Philadelphia. In: FASOLD, Ralph W.; SCHIFFRIN, Deborah (Eds.). *Language Change and Variation*. Washington, DC: Georgetown University Press, 1989b. p. 1-57.

LABOV, W. *Principles of linguistic change: internal factors*. Oxford: Blackwell, v. 1, 2001.

LABOV, W. *Principles of linguistic change: social factors*. Oxford: Blackwell, v. 2, 2001.

LABOV, W. The anatomy of style-shifting. ECKERT, Penélope; RICKFORD, John R. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 2002.

LADEFOGED, Peter. *Phonetic Data Analysis: an Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell, 2003.

LAMPRECHT, R. P. A aquisição da fonologia do Português na faixa etária dos 2:9 e 5:5. *Letras de hoje*, 28, 2, p. 95-106, 1993.

LEITE, Camila Tavares. *Seqüências de (oclusiva alveolar + sibilante alveolar) como um padrão inovador no português de Belo Horizonte*. 2006. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2006.

LE PAGE, R. B. Problems of description in multilingual communities. *Transactions of the Philological Society*, p. 189-212, 1968.

LEVELT, Clara C.; SCHILLER, Niels O.; LEVELT, William J. The acquisition of syllable types. *Language acquisition*, 8 (3), p. 237-264, 2002.

LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM (LAEL). Corpus oral e escrito da língua portuguesa. Disponível em: <<http://LAEL.pucsp.br/corpora/index.htm>>. Acesso em: fev. e mar. 2004.

MARCUSCHI, Luiz Antonio. *Da fala para a escrita: atividades de retextualização*. 4 ed. São Paulo: Cortez, 2003.

MARSHALL, Jonathan. *Language change and sociolinguistics: rething social networks*. New York: Palgrave Macmillan, 2004.

MATTOSO CÂMARA JR., Joaquim. *Estrutura da língua portuguesa*. 14 ed. Petrópolis: Vozes, 1970.

McDONALD, Scott A.; SHILLCOCK, Richard C. Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech*, v. 44, n. 3, p. 295-323, 2001.

MEILLET, Antoine. *Linguistique historique et linguistique générale*. 2 ed. Paris: Librairie Ancienne Honoré Champion, 1926.

MENON, Odete P. S.; LOREGIAN-PENKAL, Loremi. Variação no indivíduo e na comunidade: tu e você no sul do Brasil. In: VANDRESEN, Paulino (Org.) *Variação e mudança no português falado da região Sul*. Pelotas: Educat, 2002, p. 147-188.

MILROY, Lesley. *Language and Social Networks*. 2 ed. Oxford: Basil Blackwell, 1987.

MILROY, Lesley. "On the social origins of language change". (capítulo 8), 1993.

MILROY, Lesley. Social Networks. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002. p. 549-572.

MILROY, Lesley. Introduction: mobility, contact and language change – working with contemporary speech communities. *Journal in Sociolinguistics*, v. 6, n. 1, p. 3-15, 2003.

MIRANDA, Izabel. *Aquisição e variação estruturada de encontros consonantais tautossilábicos*. Belo Horizonte. 2007. Tese (Doutorado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2007.

MOHANAN, K. P. Emergence of complexity in phonological development. In: FERGUSON, C. A.; MENN, L.; STOEL – GAMMON, C. (Eds.). *Phonological Development: Models, research, implications*. Timonium, MD: York Press, 1992.

MOLLICA; BRAGA (Orgs.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003

MOTA, Helena Bolli. *Terapia Fonoaudiológica para os desvios fonológicos*. Rio de Janeiro: Revinter, 2001.

MOTA, J. A. *O <s> em coda silábica na norma culta de Salvador*. 2002. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002.

NARO, Anthony. O dinamismo das línguas. In: MOLLICA; BRAGA (Orgs.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003, p. 43-50.

OGURA, Mieko; WANG, William S-Y. Lexical diffusion and evolution theory. In: HICKEY, Raymond; PUPPEL, Stanislaw. *Language History and Linguistic Modelling*. Berlin: Mouton de Gruyter, 1997.

OLIVEIRA, Alan Jardel de. *Variação em itens lexicais terminados em /l/ + vogal na região de Itaúna/MG*. 2006. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2006.

OLIVEIRA, Marco Antonio de. The neogramarian controversy revisited. *International Journal of the sociology of language*, Berlin, v. 89, p. 93-105, 1991.

OLIVEIRA, Marco Antonio de. Aspectos da difusão lexical. *Revista de Estudos da Linguagem*. Belo Horizonte: Faculdade de Letras da UFMG, v. 1, p. 31-41, 1992.

OLIVEIRA, Marco Antonio de. O léxico como controlador das Mudanças Sonoras. *Revista de Estudos da Linguagem*. Belo Horizonte: Faculdade de Letras da UFMG, v. 1, p. 75-92, 1995.

OLIVEIRA, Marco Antonio de. Reanalizando o processo de cancelamento do (r) em final de sílaba. *Revista de estudos da linguagem*, Belo Horizonte: v. 6, n. 2, p. 31-58, jul./dez. 1997.

OLIVEIRA, Marco Antonio de; LEE, Seung Hwa. Variação Inter- e Intra-Dialetal no Português Brasileiro: Um Problema para a Teoria Fonológica. In: DA HORA, Demerval; COLLISCHONN, Gisela. (Orgs.). *Teoria Lingüística – Fonologia e Outros Temas*. 1 ed. João Pessoa: Editora Universitária – UFPB, 2003, v. 01, p. 67-91.

OLIVEIRA, Marco Antonio de. Variação Lingüística e Teoria Fonológica. In: XXI Encontro Nacional da ANPOLL, São Paulo, 2006.

OLIVEIRA GUIMARÃES, Daniela Mara Lima. *Variação nas seqüências de (sibilante +africada alveopalatal) no português de Belo Horizonte*. 2004. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2004.

OLIVEIRA GUIMARÃES, Daniela Mara Lima. *Aquisição de alofones: um estudo sobre as africadas alveopalatais no português brasileiro*. Em andamento. Tese (Doutorado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte. Em andamento.

OSTHOFF; BRUGMAN. Preface to morphological investigations in the sphere of the Indo-European languages. 1878.

PAGANO, Marcello; GAUVREAU, Kimberlee. *Princípios de Bioestatística*. Tradução da 2ª edição norte-americana. Luiz Sérgio de Castro Paiva (tradução). São Paulo: Pioneira Thomson learning, 2004.

PAGOTTO, Emílio Gozze. *Variação e (‘) identidade*. Maceió: UFAL, 2004.

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

PALMERI, T. J.; GOLDINGER, S. D.; PISONI, D. B. Episodic encoding of voice attributes and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19 (2), p. 309-328, mar. 1993.

PHILLIPS, B. Word frequency and the actuation of sound change. *Language*, v. 60, n. 2, p. 320-42, 1984.

PHILLIPS, B. Lexical Diffusion is not lexical analogy. *Word*, v. 49, n. 3, p. 369-381, 1998.

PHILLIPS, B. The mental lexicon: evidence from lexical diffusion. *Brain and language*, v. 68, p. 104-109, 1999.

PIERREHUMBERT, Janet. Stochastic phonology. *Glott International*, v. 5, n. 6, p. 195-207, 2001a.

PIERREHUMBERT, Janet. Exemplar dynamics: Word frequency, lenition, and contrast. In: BYBEE, J.; HOPPER, P. (Ed.) *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001b. p. 137-157.

PIERREHUMBERT, Janet. Word-specific phonetics. In: GUSSENHOVEN, C.; WARNER, N. (Ed.). *Laboratory Phonology VII*, Mouton De Gruyter, Berlin, 2002, p. 101-140.

PIERREHUMBERT, Janet. Probabilistic Phonology: Discrimination and Robustness. In: BOD, R.; HAY, J.; JANNEDY, S. (Ed.). *Probability Theory in Linguistics*. MIT Press, 2003a. p. 1-33. Disponível em: <[www.ling.nwu.edu/~jbp/publications.html](http://www.ling.nwu.edu/~jbp/publications.html)>. Acesso em: 28 out. 2003.

PIERREHUMBERT, Janet. Probabilistic Phonology. In: BOD, J.; HAY, J.; JANNEDY, S (Eds.) *Probabilistic Linguistics*. Cambridge: MIT Press, 2003b.

PIERREHUMBERT, Janet. *Phonological dynamics in individuals and populations*. Phonetics and Linguistics Seminars. Departament of Phonetics and Linguistics (University College London – UCL). Trabalho apresentado em outubro de 2004. Disponível em: <<http://www.phon.ucl.ac.uk/>>. Acesso em: 21 jul. 2004.

PISONI, D. Some thought on 'normalization' in speech perception. In: JOHNSON, K.; MULLENNIX, J. (Eds.). *Talker variability in speech processing*. San Diego: Academic Press, 1997. p. 9-32.

PRAAT. Programa de Análise Acústica gratuito na Internet. Disponível em: <<http://www.fon.hum.uva.nl/praat/>>. Acesso em: 11 ago. 2004.

PROJETO ASPA. Avaliação Sonora do Português Atual. Disponível em: <<http://www.projetoaspa.org/>>. Acesso em: mar. 2006.

REDFORD, Melissa; CHEN, Chun Chi; MIIKKULAINEN. Constrained Emergence of Universals. *Language and speech*, v. 44, n. 1, p. 27-56, 2001.

REINECKE, Katja. *Os róticos intervocálicos na gramática individual de falantes de Blumenau e Lages*. 2006. Tese (Doutorado em Lingüística) – Faculdade de Letras, UFSC, Florianópolis, 2006.

ROCHA, Luiz Carlos de Assis. *Estruturas morfológicas do português*. Belo Horizonte: UFMG, 1998.

ROSCH, Eleanor. Cognitive representations of semantic categories. *Journal of Experimental Psychology*, v. 104, n. 3, p. 192-233, 1975.

ROSCH, Eleanor *et al.* Basic objects in natural categories. *Cognitive Psychology* 8, p. 382-439, 1976.

RUSSELL, J. Networks and sociolinguistics variation in an African urban setting. In: Romaine (Ed.), p. 125-140, 1982.

SANGSTER, Catherine M. *Inter- and Intra-Speaker Variation in Liverpool English: A Sociophonetic Study*. 2002. 290 f. Tese de Doutorado – University of Oxford, Oxford, 2002.

SAPIR, Edward. A realidade psicológica dos fonemas. In: DASCAL, M. (Org). *Fundamentos Metodológicos da Lingüística – Fonologia e Sintaxe*. Campinas: IEL-Unicamp, v. 2, p. 37-55, 1981.

SAUSSURE, Ferdinand. *Curso de lingüística geral*. São Paulo: Cultrix, 1916.

SCHACTER, D. L.; CHURCH, B. A. Auditory priming and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18 (5), p. 915-930, set. 1992.

SELTING, Margret; COUPER-KUHLEN, Elizabeth. Argumente für die Entwicklung einer ‘interaktionalen Linguistik’. *Gesprächsforschung*. Online – Zeitschrift zur verbalen Interaktion (ISSN 1617-1837), 2000, p. 76-95. Disponível em: <[www.gespraechforschung-ozs.de](http://www.gespraechforschung-ozs.de)>.

SHEN, Z. Lexical diffusion: a population perspective and a mathematical model. *Journal of Chinese Linguistics*, [S.L], v. 18, p. 159-291, 1990 *apud* OLIVEIRA GUIMARÃES, Daniela Mara Lima. *Variação nas seqüências de (sibilante +africada*

*alveolopalatal) no português de Belo Horizonte*. 2004. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2004.

SINGLETON, David. *Language and the lexicon: an introduction*. London: Arnold, 2000.

SMITH, Jonh. Types and tokens in language change: some evidence from Romance. In: HICKEY, Raymond; PUPPEL, Stanislaw. *Language History and Linguistic Modelling*. Berlin: Mouton de Gruyter, 1997.

SOARES, Magda Becker. *Alfabetização e Letramento*. São Paulo: Contexto, 2003.

SZCZENIAK, Konrad. Palavras-relâmpago: como aprendemos e utilizamos nosso vocabulário. *Ciência Hoje*, v. 35, n. 207, p. 16-20, ago. 2004.

THELEN, E.; SMITH, L. B. *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press, 1994.

TOMAZ, Kátia Silva. *Alternância de vogais médias posteriores em formas nominais de plural no português de Belo Horizonte*. 2006. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, UFMG, Belo Horizonte, 2004.

VIEGAS, Maria do Carmo. Estudo sobre o alçamento de vogais médias pretônicas e os itens lexicais. *Revista do CESP*, Belo Horizonte, v. 21, n. 28/29, p. 217-241, jan.-dez. 2001.

VIHMAN, Marilyn May. *Phonological Development: the origins of language in the child*. Cambridge: Blackwell, 1996, 312 p.

YAVAS, M. Padrões na aquisição da fonologia do Português. *Letras de hoje*, 23, p. 7-30, 1988.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. Empirical foundations for a theory of language change. In: LEHMANN, W. P.; MALKIEL, Yakov (Eds.). *Directions in Historical Linguistics*. Austin, TX:University of Texas Press, 1968. p. 95-188.

WILDGEN. Das dynamische Paradigma in der Linguistik. 1987, republicado em 2005. Disponível em: <[www.fb10.uni-bremen.de/homepages/wildgen/pdf/das\\_dynamische\\_paradigma.pdf](http://www.fb10.uni-bremen.de/homepages/wildgen/pdf/das_dynamische_paradigma.pdf)>.

## APÊNDICE A

### Traduções nossas para as citações em língua inglesa

A seguir, são apresentadas as traduções nossas para as citações em língua inglesa, conforme a ordem em que estas ocorreram nesta tese.

#### **Página 19**

Hazen (2002, p. 501, 502) → “Para alguns pesquisadores, padrões de variação da língua de indivíduos são idioletos. Contudo, de acordo com a análise variacionista tradicional, especificamente Weinreich et al. (1968) e Labov (1989b), idioletos não são uma realidade teórica. [...] Nesta visão, a gramática da variação lingüística existe no nível da comunidade de fala, mas não no nível do indivíduo. Essas idéias impedem uma visão aninhada de variação lingüística, segundo a qual dialetos são coleções de idioletos similares, e línguas são coleções de dialetos similares.”

#### **Página 20**

Milroy (1987, p. 132, 133) → “Labov (1966) percebe considerável variação dentro de grupos sociais definidos com base em idade, classe ou sexo, mas argumenta, convincentemente, que estudos de idioletos são improváveis de revelar estrutura sociolingüística tão claramente quanto estudos da fala de grupos sociais inteiros. [...] Notamos aqui que Labov não está sugerindo que indivíduos revelam somente variações desestruturadas; ele está mostrando, especificamente, que a abordagem baseada no idioleto, característica de estudiosos anteriores (por exemplo, Bloch, 1948; Hockett, 1958), é inadequada”.

#### **Página 21**

Marshall (2004, p. 26) → “A prática de agrupar informantes para mostrar padrões de variação lingüística é, claro, um procedimento válido; contudo, muita variação individual sistemática é desprezada por esse método”

Milroy (1987, p. 133) → “Entretanto, já que não esperamos encontrar na língua do indivíduo uma réplica perfeita em miniatura da sua gramática dialetal, não há razão porque a produção de um único falante deva ser vista com desestruturada e imprópria de ser estudada. De fato, Bickerton, Le Page e Gumperz [Bickerton (1975), Le Page (1968) e Gumperz (1976b)], todos tomam o indivíduo como a base de suas descrições de variação, principalmente por causa da dificuldade de se definir limites de grupos. Le Page, adicionalmente, tem a preocupação de enfatizar a liberdade de escolha do indivíduo de usar sua língua variavelmente como um significado de identificação com diferentes grupos em diferentes tempos”.

#### **Página 22**

Docherty *et al.* (1997, p. 276) → “[...] padrões de variação (ambos dentro dos e entre os indivíduos) emergem um tanto sistematicamente, mas são mais complexos e de maior magnitude do que tem sido sugerido na literatura fonológica”.

#### **Página 35**

Shen (1990, p. 166) → “É muito provável que uma palavra diferente vai mudar sua categoria fonológica de A para B em um tempo totalmente diferente para os seus aprendizes. O que é observado é que irregularidade lexical aparece. Ao longo da dimensão social, por causa de vários ambientes da língua, enquanto um indivíduo mudou a identidade fonológica de uma palavra, ao mesmo tempo, outros podem ainda manter o original”.

#### **Página 37**

Ash (2002, p. 402) → “Classe social é um conceito central na pesquisa sociolingüística.”

**Página 38**

Milroy (1987, p. 137) → “Ainda, apesar dessas diferenças lingüísticas que podem ser ligadas firmemente às variáveis sexo, idade e origem geográfica, há um grande resíduo de variação sistemática entre indivíduos que não podem ser caracterizados de um modo claro pela divisão de falantes em mais subgrupos”.

**Página 39**

Milroy (1987, p. 135) → “[...] o conceito de rede foi desenvolvido com o verdadeiro propósito que requeremos agora – explicar tipos de comportamento individual que não podem ser explicados em termos de membro de grupo”.

**Página 40**

Labov (2001, v. 2, p. 410) → “A história de nossos líderes da mudança lingüística é uma história de não-conformidade, e sua posição sociolingüística é uma mostra de não-conformidade”.

Labov (2001, v. 2, p. 410) → “[...] negociação, persuasão e denúncia, todos inscritos sob uma profunda intolerância para “cuspidity”, hipocrisia e injustiça. Essas são as qualidades que fazem um grande líder de mudança lingüística”.

**Página 47**

Bybee (2001, p. 1) → “Este livro introduz, ao estudo tradicional de fonologia, a noção de que o uso da língua desempenha um papel no ajuste da forma e do conteúdo dos sistemas sonoros”.

**Página 54**

Vihman (1996, p. 5) → “[...] não há, no meu conhecimento, evidência de que falantes adultos de uma língua dividam uma gramática idêntica, apesar das hipóteses nativistas (ver novamente Mohanan, 1992). Ao contrário, diferenças individuais são exibidas pelos adultos [...]”.

**Página 61**

Pierrehumbert (2001b, p. 139) → “A capacidade da mente para memórias de longo termo de exemplos individuais é, de fato, surpreendentemente grande, como experimentos vistos em Johnson (1997) indicam”.

**Página 62**

Pierrehumbert (2001b, p. 142) → “A Teoria de Exemplos provê-nos um modo de formalizar o conhecimento fonético detalhado que falantes nativos têm sobre as categorias de sua língua. Visto que a teoria de exemplos armazena diretamente a distribuição de valores de parâmetro fonético associados com cada marca, isto provê-nos uma imagem do “conhecimento fonético implícito do falante”.

**Página 63**

Pierrehumbert (2001b, p. 139) → “O volume de fala que uma pessoa processa em um tempo de vida é tão extenso que seria indesejável assumir memórias individuais de cada uso de cada palavra”.

**Página 65**

Bod; Hay; Jannedy (2003, p. vii) → “a linguagem é probabilística”.

Bod; Hay; Jannedy (2003, p. 7) → “Probabilidades permeiam o sistema lingüístico”.

Bod; Hay; Jannedy (2003, p. vii) → “Na fonologia [...] probabilidades desempenham um papel em julgamentos de aceitabilidade e alternâncias”.

Bod; Hay; Jannedy (2003, p. 1) → “A Lingüística Probabilística amplia o conjunto de dados, o qual uma teoria pode explicar e pelo qual deve ser responsável”.

Pierrehumbert (2003b, p. 226) → “[...] entidades de todos os níveis de representação na fonética e na fonologia demonstram variação estatística”.

### **Página 66**

Bod; Hay; Jannedy (2003, p. 4) → “Categorias são centrais na teoria lingüística, mas membros nessas categorias não precisam ser categóricos. A Lingüística Probabilística conceitualiza categorias como distribuições. Membro nas categorias é gradiente”.

Bod; Hay; Jannedy (2003, p. vii) → “A abordagem probabilística para o estudo da linguagem pode parecer oposta à abordagem categórica, que tem dominado a lingüística por mais de 40 anos. [...] enquanto abordagens categóricas focalizam o ponto final das distribuições dos fenômenos lingüísticos, abordagens probabilísticas focalizam o terreno meio gradiente”.

Bod; Hay; Jannedy (2003, p. 10) → “A língua mostra todas as marcas de um sistema probabilístico. Categorias e boa formação são gradientes, e efeitos de frequência estão em toda parte. Acreditamos que todas as evidências apontam para uma faculdade de linguagem probabilística. O conhecimento da língua deve ser entendido não como um conjunto mínimo de regras ou restrições categóricas, mas como um (possivelmente redundante) conjunto de regras gradientes, que podem ser categorizadas por uma distribuição estatística”.

Bod; Hay; Jannedy (2003, p. 2, 3) → “Dentro de indivíduos, padrões de produção diferem com base em fatores estilísticos, tais como, destinatário, contexto e tópico, e essa variação estilística de grande grau ecoa a variação presente nos membros da sociedade. [...] E indivíduos diferem não somente nas variantes específicas que eles usam em diferentes contextos, mas também na frequência com que eles as usam. Conhecimento de variação deve envolver conhecimento de frequências”.

### **Página 67**

Bod; Hay; Jannedy (2003, p. 3) → “Um surpreendente indício da importância das probabilidades na língua vem da riqueza dos efeitos de frequência que permeiam: representação lingüística, processamento e mudança lingüística.

[...] Podemos melhor modelar muitos desses efeitos ao fazer explícita a ligação entre frequência e probabilidade. A teoria da probabilidade proporciona métodos bem articulados para modelar frequência [...].”

### **Página 68**

Bod; Hay; Jannedy (2003, p. 3, 4) → “Somente a completa ausência de efeitos de frequência desafiaria as fundações da lingüística probabilística, então, sua esmagadora presença adiciona peso à reivindicação de que a faculdade de linguagem é inerentemente probabilística.

[...] Efeitos de frequência proporcionam um tipo de evidência para uma lingüística probabilística.”

## APÊNDICE B

A seguir, no roteiro da entrevista de fala espontânea utilizado nesta pesquisa, as 30 palavras em estudo estão em negrito.

### Roteiro da entrevista de fala espontânea com as 30 palavras em estudo

- 1) Você usa **computador**? Gosta do **computador**? Acha o **computador** importante? Se não usa, gostaria de usar, de ter um **computador**? O que você acha da presença do **computador** hoje no mundo? Você pensa que o **computador** veio para ficar? Você acha que, com o desenvolvimento do **computador**, o trabalho braçal vai acabar?
- 2) Você acha que **qualquer** companhia é melhor do que nenhuma? Você concorda com a famosa frase: “**qualquer** maneira de amor vale a pena”?
- 3) Quem para você é um **grande** artista? Um **grande** cantor, ator...? Você gosta de família **grande**? E de casa **grande**?
- 4) Quando você tem algum **problema**, o que você faz para se livrar dele? Você lida bem com uma situação **problema**? Você concorda com a frase: “**problema** todo mundo tem”? Você acha que **qualquer** um pode ser político profissional? Qual você acha que é o maior **problema** que o **Brasil** tem?
- 5) **Dinheiro** é tudo na vida? Você acha que o **dinheiro** é o mais importante na vida? O que você acha de pessoas que só pensam em **dinheiro**? Coloque em ordem de importância essas **quatro** coisas: **dinheiro**, família, lazer e relação amorosa (O quê que vem em **primeiro**, em segundo, em terceiro e quarto **lugar**?).
- 6) Você acha que algum dia o real vai alcançar o **dólar**? Você acha que a cotação do **dólar** influencia a nossa moeda, o real?
- 7) Você acha certo um **jogador** de futebol ganhar tanto **dinheiro**, muito mais do que muitos outros profissionais que estudaram e estudam a vida inteira? Na sua opinião, qual o melhor **jogador** do **Brasil**? E do mundo? O que você pensa do Ronaldinho como **jogador**? Você acha que ele é tão bom **jogador** como a mídia propaga?
- 8) Você gosta de ir a alguma **feira**? Em Belo Horizonte, **para** você qual é a **melhor feira**? Qual é o **melhor** dia da semana **para** trabalhar **para** você? E o pior? (espera-se que o informante responda a estas duas últimas perguntas: segunda-**feira**, terça-**feira**).
- 9) Você gosta de **mar**? Muitos acham que o **mar** é perigoso, você pensa isso também? O que você prefere **mar** ou campo, roça?
- 10) Quais são as **três** coisas que você mais gosta de fazer? Quais as **três** coisas mais importantes na vida **para** você? Se aparecesse um gênio na sua frente e você pudesse pedir **três** coisas, o que você pediria? Quantos eram os reis magos?
- 11) Você acha que a corrupção no **Brasil** pode acabar (**ela**)? Você acha que o **Brasil** é um país bom para se viver? Por quê? Você gosta de viver no **Brasil**? Se você pudesse escolher um **outro lugar**, país **para** viver, qual você escolheria? Por quê? O que você acha que o **Brasil** precisa para melhorar?
- 12) O que você pensa do nosso **Presidente Lula**? Como você acha que um **presidente** deve ser? Muitos afirmam que o **Lula** não ganhou nas eleições para **presidente** antes, porque tinha uma imagem negativa, a de um sindicalista arruaceiro. Você acha que um candidato a **presidente** de um país deve ter mesmo atenção com a sua aparência? Você acha que o governo do **Presidente Lula** está melhorando ou vai melhorar o nosso país? Você acha que os problemas de um país são decorrentes da má gestão de seu **presidente**?
- 13) O que você pensa do nosso **Governador** Aécio Neves? Muitos afirmam, inclusive cientistas políticos, que o Aécio Neves tem a intenção de passar de **Governador** de Minas para **Presidente** do **Brasil**, você acha que ele seria um bom **presidente** para o nosso país, considerando como ele é como **governador**? Você acha que ele irá conseguir ser **presidente** de nosso país? O que você achou do Itamar Franco como **governador**? Ele foi muito polêmico como **governador**, mas o que você achou dele como **presidente** do **Brasil**?
- 14) O que você pensa dos políticos **Brasil** (**eles**)?
- 15) Você acha que há certas mágoas que ficam na pessoa **para sempre**? Duas pessoas podem se amar **para sempre**? Você acredita o amor entre um homem e uma mulher em um casamento pode durar **para sempre**? Que amor **para** você dura **para sempre**?

- 16) Você usa **óculos**? Você gosta de usar **óculos** escuros? Você acha que os **óculos** são um acessório bonito? Você acha que as pessoas ficam bem de **óculos**? E você?
- 17) Qual o **lugar** – cidade, estado, país – que você já foi que mais gostou? Qual **lugar** de sua casa onde você mais gosta de ficar? Você está aqui me cedendo essa entrevista gentilmente, mas se você pudesse, **agora**, em que **lugar** você gostaria de estar? Qual é **para** você o **melhor lugar** para se descansar? Por quê?
- 18) Como foi o seu **primeiro** ano na escola? Você acha que o **primeiro** amor é mesmo marcante na vida de uma pessoa? O que você acha que deve vir **primeiro** na vida: a carreira ou o casamento?
- 19) Você acha que a família serve **para** que na vida da gente?
- 20) *Quantas são as estações do ano?* Resposta esperada: **quatro**. Você gosta de todas as **quatro** estações do ano? Qual das **quatro** você gosta mais?
- 21) Qual é a **melhor** hora do dia para você? (ou: qual é o seu(sua) **melhor** amigo(a)? Por que ele(a) é o seu(sua) **melhor** amigo(a)? Qual é ou foi a **melhor** fase da sua vida? Porquê? Qual é a **melhor** fase da vida do ser humano? Qual a **melhor** lembrança que você tem?)
- 22) Qual foi a **maior** emoção que você já sentiu até hoje? Qual é o seu **maior** sonho? Você acha que uma pessoa de 16 anos poderia ser considerada **maior** de idade pela lei? Qual a faixa **etária** mais adequada para uma pessoa ser considerada **maior** de idade, na sua opinião? Quem foi o **maior** exemplo na sua vida? Quem, para você, é o **maior** exemplo de vida? Qual foi a sua **maior** alegria até hoje? Por quê?
- 23) Você prefere a vida que você vive **agora** ou você gostaria de voltar ao passado? Se você pudesse escolher um **lugar** para estar, onde você estaria **agora**? Você acha que a questão do desemprego no **Brasil** deve ser resolvida **agora** ou você acha que dá para esperar mais algum tempo?
- 24) O que você mais gostava de fazer quando **era** criança? Você acha que o **Brasil** de uns anos atrás **era melhor** do que o **Brasil** de hoje?
- 25) Você gostaria de morar em **outro lugar**? Você gostaria de ter **outro** nome? Se você pudesse escolher um **outro** nome, que **outro** nome você escolheria? Se você pudesse escolher um **outro lugar**, que **outro lugar** você escolheria?
- 26) Você gosta de ler **livro**? Qual o **livro** que você que você mais gostou? Há algum **livro** que você ainda quer ler? Qual? Você acha que ler **livro** é importante para o crescimento cultural de uma pessoa? Há algum **livro** que você leu que você não gostou?
- 27) Qual a **mulher** que você mais admira (uma **mulher** próxima de você, alguém famoso)? Por quê? (ou: o que você acha da postura da **mulher** na sociedade hoje?)

## APÊNDICE C

### Tarefa de leitura

#### **As contadoras de histórias de Tempo Feliz**

Dona Glória e Dona Marta, duas irmãs gêmeas de oitenta e seis anos, são contadoras de histórias em um lugar pacato no norte de Minas chamado Tempo Feliz. Em geral essas irmãs passam a maior parcela do seu tempo treinando a dramatização das narrativas. Elas dizem que essa é a melhor tarefa que fazem na vida. Na opinião de todos, os contos são sempre perfeitos e têm um enredo muito envolvente. Antes de serem contadoras de histórias, a Dona Glória era tenista e a Dona Marta era protética e também dentista. Mas as duas já tiveram outros empregos antes ainda, sendo que o primeiro trabalho de Dona Glória foi de secretária e o primeiro trabalho de Dona Marta foi de professora. Depois de aposentarem é que elas descobriram esse dom delas de contar histórias. As duas irmãs têm o maior talento para contar “causos”, o que é comprovado pelo fato de que em toda e qualquer parada delas em algum lugar parece haver uma festa entre as pessoas que ouvem as histórias que elas contam.

O público é que escolhe o que elas vão contar. Dona Glória e Dona Marta atendem geralmente a três pedidos de histórias em cada sessão. As sessões ocorrem onde elas moram na Rua Quatro Paredes pelo menos duas tardes por semana. Todos ficam encantados com as histórias principalmente porque o grande conjunto delas são “causos” conhecidos pelas pessoas da cidade. Mas as duas irmãs às vezes contam coisas que acontecem agora também no nosso Brasil tachado de país de terceiro mundo por causa do problema político-econômico e social que está sempre perante a nossa população sofredora. Por exemplo, elas já estão tratando do Mensalão que é considerado como o problema político mais sério que já ocorreu nesse nosso grande país de gente humilde mas trabalhadora.

Só que Dona Glória e Dona Marta inventam coisas em cima desses acontecimentos atuais. Elas contam uma história em que até um jogador parrudo de futebol está envolvido no caso do Mensalão. De acordo com elas, o que acontece é que esse tal jogador parrudo atua muito bem em um dia mas muito mal em outro porque faz o placar que o político quiser em qualquer partida em troco de um Mensalão.

As irmãs não poupam nem mesmo o Lula contando que esse Presidente casado com o compromisso de fazer o bem à nossa nação supostamente e o Governador calado e indiferente do Acre estão envolvidos no caso do Mensalão. Nessa história elas contam que o Lula concede ilegalmente um alto valor em dinheiro parado na conta da presidência ao tal Governador também do partido desse Presidente querido pelo povo por ter sido um operário que agora pertence ao pequeno grupo que dirige o país. Dona Glória e Dona Marta criticam esses políticos dizendo que eles teriam de devolver essa quantia ao povo em forma de emprego. Mas quando as irmãs terminam essa história, elas se desculpam com eles pedindo ao público que não leve a sério a história delas que é inventada. O público aplaude entusiasmado. Todo mundo da cidade de Tempo Feliz e do Brasil parece estar abismado com esse caso do Mensalão e é por isso que ele talvez tem esse ibope enorme entre as pessoas que ouvem as histórias das irmãs.

Dona Glória e Dona Marta ainda escrevem e declamam poemas. Elas têm até um poema moderno que fala de um computador capaz de dominar os homens. A história desse computador capaz de mandar nos homens provoca muita risada entre as pessoas e alguns pensam que ela permite pensar que é possível acontecer o dia em que a máquina vai substituir o homem. Mas ainda assim nenhum outro poema é mais interessante do que o que fala do caso da mulher pagã que trabalha na Feira Cativa perto do Mar Tainha vendendo um livro pesado que se chama “O óculos preto que faz enxergar o futuro”. O mais engraçado é que as irmãs contam que esse livro pesado que fala de um óculos preto que consegue fazer previsões do futuro é vendido em dólar parado em conta bancária. Segundo elas, se o cliente não tiver dólar parado na conta bancária não é possível vender o produto mesmo porque é só pela internet que essa mulher pagã realiza a venda. Assim não adianta o freguês ir na Feira Cativa próxima ao Mar Tainha com dinheiro querendo comprar a obra. A feirante só vende esse famoso produto pela internet. O detalhe é que ela também vende três produtos apenas por freguês. Uma pessoa que quis levar quatro produtos teve de pedir ao seu melhor parente e amigo para comprar a obra como se ele quisesse comprar e não essa pessoa. Quem quiser saber mais detalhes sobre essa e outras histórias, procure Dona Glória e Dona Marta, contadoras de histórias da cidade de Tempo Feliz.

## APÊNDICE D

### **Anotações a respeito das coletas e dos 2 participantes do estudo longitudinal**

Abaixo, além de serem feitas algumas observações sobre os 2 participantes do estudo longitudinal, é apresentada uma breve descrição de cada uma das seis coletas feitas com tais informantes: primeiro, para a informante líder L2R e depois, para o informante não-líder N3Z.

#### **Informante líder L2R –proveniente do Corpus 2 (FONTES MARTINS, 2001) – e suas 6 coletas**

Informante feminino

Idade: 26 anos

Naturalidade: Belo Horizonte

Escolaridade: superior, formada em Ciências Contábeis

Profissão: Contadora

Solteira, mora com os pais

Características da fala da entrevistada: fala bem mineira, fala rápido.

As gravações foram feitas de outubro de 2005 a março de 2006, de mês a mês, num total de 6 gravações/coletas de, aproximadamente, 30 minutos cada uma. Assim, o tempo total de duração das gravações é de aproximadamente 3 horas.

Das 6 coletas, 4 foram realizadas na casa da informante e 2, em um restaurante próximo ao seu trabalho, no seu horário de almoço. As gravações na casa dela foram feitas à noite, após o seu horário de trabalho, visto que este era o único tempo de que a informante dispunha. Os pais dela estavam sempre em casa e faziam questão de receber a pesquisadora. Já as gravações no restaurante foram feitas no segundo andar deste, onde havia poucas pessoas e menos ruído. Além disso, a pesquisadora e a informante se sentaram em uma mesa distante das outras pessoas, para evitar ruído.

A informante trabalha em horário integral, na área de contabilidade de uma grande empresa de engenharia. Ela ainda estuda: faz pós-graduação em dois dias da semana à noite e aos sábados. Com isso, ela dispõe de pouco tempo: restam-lhe somente os domingos para descansar realmente.

A informante é muito apegada à sua família. Aliás, esse é um dos temas preferidos dela, além dos amigos de faculdade, do trabalho e do atual estudo. Ela é muito aplicada, gosta muito do seu trabalho, dizendo-se uma pessoa do “mundo dos números”. Gosta muito de estudar também e vem se preparando para concursos públicos na sua área.

Assim, na primeira coleta em sua casa, a informante falou mais sobre seu estudo atual (a pós-graduação e a preparação para concursos públicos) e sobre os seus tempos e amigos da faculdade de ciências contábeis, até mesmo porque esse era um assunto que a aproximava da pesquisadora, já que esta é irmã de uma colega de faculdade da informante.

Nas segunda e terceira coletas também em sua casa, a informante falou muito sobre a sua formatura no ano de 2004, da qual ela foi chefe da comissão de formatura. Ela contou vários detalhes de como organizou os eventos (colação de grau, missa, baile) e uma parte dessas entrevistas foi baseada em fotos: ela mostrava e comentava as fotos da formatura. Ela falou ainda sobre os amigos de curso, o que eles estão fazendo depois de formados.

Na quarta coleta no restaurante, ela falou sobre o que faz em seu tempo livre (assiste a filmes, visita parentes...) e os lugares, especialmente restaurantes e bares, onde gosta de ir. Na quinta coleta em sua casa novamente, ela falou mais sobre sua família, sobre o trabalho de sua mãe e do seu pai que hoje é aposentado, sobre comidas de que gosta e sobre questões econômicas do Brasil. Na sexta e última coleta no restaurante, a informante falou sobre o trabalho e os estudos novamente.

**Informante não-líder N3Z – proveniente do Corpus 3 (CRISTÓFARO-SILVA, 2003) – e suas 6 coletas**

Informante masculino

Idade: 56 anos

Naturalidade: Belo Horizonte

Escolaridade: superior incompleto

Profissão: trabalha na área administrativa de uma empresa

Divorciado, 2 filhos, mora sozinho.

Características da fala do entrevistado: fala devagar, articulando bem os sons.

As gravações foram feitas de outubro de 2005 a março de 2006, de mês a mês, num total de 6 gravações/coletas de, aproximadamente, 30 minutos cada uma. Assim, o tempo total de duração das gravações é de aproximadamente 3 horas.

O informante, em todas as seis coletas longitudinais, recebeu a pesquisadora em sua casa (ele é quem preferiu que as coletas fossem sempre em sua casa) com muita gentileza, educação, simpatia e atenção. As conversas foram bem espontâneas, de modo que era o informante que as conduzia mais: a pesquisadora deixava-o falar sobre o assunto que quisesse (música, amigos, família...) e como quisesse (por meio de fotos e até, às vezes, lendo textos). Em todas as coletas, ele se mostrou com maior inclinação para falar sobre coisas do passado em geral, música, fotografia, viagens, família e amigos.

O assunto principal da primeira coleta longitudinal foi a viagem do informante ao Chile. A conversa foi toda baseada nas fotos pessoais dele (aliás, outras coletas também tiveram, entre outros, fotos como base para a conversa). Ele é quem sugeriu que conversássemos sobre essas fotos inclusive.

Na segunda coleta, o assunto principal foi música, a grande paixão do informante (juntamente com a fotografia, o que ele mesmo revelou). Ele gosta, em especial, de cantores mais antigos, como, Vinícius de Moraes, Pixinguinha e Adoniran Barbosa. O informante chegou a cantar trechos de músicas, como, “Iracema” e “Saudosa Maloca” (Adoniran Barbosa).

Na terceira coleta, os assuntos foram mais voltados para as coisas antigas de modo geral (música e até a linguagem antiga). Ele se mostrou saudosos dos tempos antigos: dos seus tempos de menino, da Belo Horizonte antiga etc.

Na quarta coleta, houve uma sessão de fotos pessoais novamente: fotos da família, das viagens (Milho Verde, Serra do Cipó), dos aniversários, dos amigos.

A quinta coleta foi marcada pelo estilo leitura algumas vezes. Isso ocorreu por que o informante quis ler trechos de um texto com português escrito antigo (*Compêndio da Bayer*) e um livro com fotos antigas, em preto e branco, contando a história do Brasil. O informante se mostrou muito empolgado com esses textos antigos, aliás, ele sempre se mostrava muito interessado em coisas antigas. Na transcrição dessa quinta coleta, todos os trechos que ele leu foram transcritos entre aspas para mostrar que não se tratava de uma produção oral espontânea dele, mas de uma oralização de texto escrito ou da leitura de textos de outrem.

A sexta e última coleta teve como assunto principal, no início, um grande amigo do informante que havia morrido há uns 3 anos. Como o informante é muito afetivo, muito ligado à sua família e a seus amigos, ele se mostrou ressentido com a morte do amigo e, por isso, apresentou um tom melancólico enquanto falava desse amigo. Depois o informante passou a falar de uma viagem que fez para a cidade dos seus pais, no interior do MG, com seus irmãos e, a partir desse momento, a entrevista tornou-se mais alegre e leve, já que o próprio informante classificou essa viagem como uma das melhores que ele já fez na vida.

## APÊNDICE E

### Critérios de seleção das 50 palavras inicialmente consideradas neste trabalho

#### 1. Apresentação de dados de 3 corpora – 3 fenômenos do português brasileiro

**TABELA 1**  
As 30 palavras em análise distribuídas nos 3 corpora

Corpora	Grau de frequência	Palavras	Frequência	Números absolutos e Percentuais de cancelamento
<b>CORPUS 1:</b> <b>Huback (2003)</b>	Mais frequentes	1. Melhor	93	24 (26%)
		2. Lugar	89	35 (39%)
		3. Mulher	74	21 (28%)
		4. Qualquer	40	28 (70%)
		5. Computador	35	16 (46%)
	Menos frequentes	6. Ar	6	0 (0%)
		7. Câncer	4	0 (0%)
		8. Doutor	4	0 (0%)
		9. Familiar	4	0 (0%)
		10. Lar	4	0 (0%)
SUBTOTAL		<b>10 palavras</b>		
<b>CORPUS 2:</b> <b>Fontes Martins (2001)</b>	Mais frequentes	11. Para	859	172 (20%)
		12. Ele	214	26 (12%)
		13. Ela	169	34 (20%)
		14. Eles	144	42 (29%)
		15. Óculos	40	32 (80%)
	Menos frequentes	16. Cérebro	5	0 (0%)
		17. Etária	5	0 (0%)
		18. Silicene	5	0 (0%)
		19. Diferenciação	4	0 (0%)
		20. Miséria	4	0 (0%)
SUBTOTAL		<b>10 palavras</b>		
<b>CORPUS 3:</b> <b>Cristóforo Silva (2003a)</b>	Mais frequentes	21. Problema	77	26 (34%)
		22. Outro	58	44 (76%)
		23. Quatro	57	35 (61%)
		24. Sempre	51	28 (55%)
		25. Primeiro	48	15 (31%)
	Menos frequentes	26. Depressão	5	0 (0%)
		27. Tranquila	5	0 (0%)
		28. Bronzeada	4	0 (0%)
		29. Complicada	4	0 (0%)
		30. Traseira	4	0 (0%)
SUBTOTAL		<b>10 palavras</b>		
TOTAL GERAL		<b>30 palavras</b>		

Fonte: Huback (2003), Fontes Martins (2001), Cristóforo Silva (2003a)

A tabela 1 acima apresenta as 30 palavras que, juntamente com outras 20 a serem apresentadas na seção 1.3, serão consideradas neste trabalho. Essas 30 palavras estão distribuídas nos três corpora indicados na primeira coluna de tal tabela. A segunda coluna discrimina o grau de frequência (mais ou menos frequente) das 30 palavras que são apresentadas na terceira coluna. Os critérios para a seleção dessas palavras serão explicitados na próxima seção. A quarta coluna mostra a frequência de ocorrência de cada uma dessas palavras, e a quinta coluna apresenta os números absolutos e percentuais de cancelamento das palavras no seu respectivo corpus.

As palavras mais e menos frequentes do Corpus 1, o qual focalizou o fenômeno do cancelamento de “r” final, foram selecionadas de Huback (2003). As palavras mais e menos frequentes do Corpus 2, que enfocou o cancelamento de líquidas intervocálicas, foram selecionadas de Fontes Martins (2001). Já as palavras mais e menos frequentes do Corpus 3, que se centrou no fenômeno de cancelamento de líquidas em encontro consonantal tautossilábico, foram selecionadas de Cristóforo Silva (2003a).

#### 1.1 Critérios de seleção das palavras da tabela 1

É importante fazer aqui alguns esclarecimentos a respeito da seleção das 5 palavras mais frequentes e das 5 palavras menos frequentes de cada um dos corpora 1, 2 e 3 (Cf. TAB. 1 da seção 1). Em alguns momentos, foi difícil escolher tais palavras nesses corpora. Um exemplo de dificuldade nessa seleção pode ser apontado quando da escolha das 5

palavras menos freqüentes do Corpus 1. Há 12 palavras com uma baixa freqüência de 4 ocorrências. Assim, uma questão pode ser levantada: como escolher apenas 5 palavras dentre 12 palavras? Que critério(s) adotar para isso?

Esta é apenas uma das dificuldades com que nos deparamos na escolha das 5 palavras mais freqüentes e das 5 palavras menos freqüentes dos corpora 1, 2 e 3. Para resolver esse e outros impasses metodológicos, tivemos de estabelecer alguns critérios que passam a ser descritos.

De modo geral, no caso da escolha das palavras mais freqüentes dos corpora 1, 2 e 3, decidimos primeiramente considerar, como um critério para indicar maior freqüência, palavras que tivessem, ao mesmo tempo, maior ocorrência e maior aplicação do fenômeno analisado. O inverso ocorreu na escolha das palavras menos freqüentes dos corpora 1, 2 e 3: consideramos, como um critério para demonstrar menor freqüência, palavras que tivessem, simultaneamente, menor ocorrência e menor aplicação do fenômeno observado.

Um outro critério complementar adotado quando necessário – em especial, na seleção das palavras menos freqüentes de cada corpora – foi considerar palavras com maior propensão estrutural à aplicação do fenômeno em análise. Vale mencionar que esse critério teve de ser usado na seleção das outras 20 palavras a serem consideradas nesta pesquisa, o que será visto adiante, na seção 1.3.

Em relação ao ambiente estrutural mais favorável à realização do cancelamento de “r” final em nominais (Corpus 1), Huback (2003, p.123) afirma que: “A princípio, parece que os sufixos *-dor* e *-tor*, independentemente de serem agentivos, estão liderando o processo de mudança (...)”.

Já no caso do cancelamento de líquidas intervocálicas (Corpus 2), a posição postônica é o contexto fonético favorecedor. Como afirma Fontes Martins (2001, p. 74): “A posição postônica final pode ser apontada como um fator favorecedor à aplicação do fenômeno em estudo.” Fontes Martins (2001, p. 106) complementa: “podemos dizer que o contexto postônico final é aquele em que se observa mais recorrentemente o cancelamento de líquidas intervocálicas.”

Também quanto ao cancelamento de líquidas em encontro consonantal tautossilábico (Corpus 3), a posição postônica favorece mais a aplicação desse fenômeno, (Cf. CRISTÓFARO SILVA, 2003a). Contudo, a posição pretônica também se mostra relevante nesse caso. Isso aponta para o fato de que o ambiente átono, de modo geral, favorece o cancelamento de líquidas em encontro consonantal tautossilábico (Cf. CRISTÓFARO SILVA, 2003b).

É bom ressaltar que os três corpora (1, 2 e 3) considerados neste trabalho abordam fenômenos diferentes com metodologias diferentes. Por esse motivo, apesar de os critérios aqui adotados na seleção das palavras mais e menos freqüentes terem sido os mesmos para os 3 corpora – como se explicitou acima, ainda nesta seção – os números encontrados para as palavras dos 3 corpora, por diversas vezes, são muito distantes.

Um exemplo disso pode ser visto na tabela 1 da seção 1. De modo geral, os corpora 1 e 3 se aproximam por apresentarem valores de ocorrência parecidos para as palavras mais freqüentes. Já o Corpus 2 se distancia desses outros dois corpora, porque os valores de ocorrência que ele apresenta para as palavras mais freqüentes são bem maiores que os dos corpora 1 e 3. Desse modo, tanto na análise das palavras mais freqüentes quanto na análise das palavras menos freqüentes aqui consideradas (Cf. TAB. 1 da seção 1), é preciso relativizar os valores (valor numérico e valor percentual) de cada palavra ao seu respectivo corpus (1, 2 ou 3).

### 1.1.1 Seleção das 5 palavras mais freqüentes de cada um dos corpora 1, 2 e 3

Na escolha das 5 palavras mais freqüentes do Corpus 1, consideramos palavras que tiveram de 31 a 100 ocorrências. Dentro desse universo, selecionamos as palavras *melhor*, *lugar*, *mulher*, *qualquer* e *computador*, visto que elas apresentaram, simultaneamente, acima de 35 ocorrências e acima de 26% de aplicação do fenômeno de cancelamento do “r” final (Cf. TAB. 1 da seção 1).

Assim, no caso do Corpus 1, o critério para a seleção de apenas 5 palavras mais freqüentes foi baseado nestes números: mais de 35 ocorrências e, ao mesmo tempo, mais de 26% de cancelamento. Tais números não puderam ser adotados para os outros dois corpora, já que, como se disse na seção anterior, por várias vezes, os valores encontrados em cada um dos corpora (1, 2 e 3) se mostraram muito distantes.

No Corpus 2, a escolha das 5 palavras mais freqüentes foi baseada em Fontes Martins (2001, p. 103) que, ao explicitar o critério de freqüência de que se utiliza, cita quais foram as 5 palavras mais freqüentes em seu corpus:

Selecionamos essas palavras [mais expressivas estatisticamente] segundo o critério de freqüência – no sentido de elas terem ocorrido mais em nosso corpus, acima de 2% de ocorrência – e também pelo maior índice de cancelamento de líquidas intervocálicas em tais palavras. Estes itens são: *ela*, *ele*, *eles*, *óculos* e *para*. (FONTES MARTINS, 2001, p. 103).

Como se pode perceber por meio dessa citação, o critério que Fontes Martins adota para encontrar as palavras mais freqüentes de seu corpus é o mesmo que adotamos aqui (Cf. seção 1.1). Maiores detalhes sobre os dados quantitativos dessas palavras – *ela*, *ele*, *eles*, *óculos* e *para* – já apresentadas na tabela 1 da seção 1, podem ser conferidos na tabela 2.

Quanto ao Corpus 3, as cinco palavras mais frequentes escolhidas foram: *problema, outro, quatro, sempre e primeiro* (Cf. TAB. 1 da seção 1). Neste corpus, consideramos como mais frequentes palavras que apresentaram, concomitantemente, mais de 48 ocorrências e acima de 30% de aplicação do fenômeno de cancelamento de líquidas em encontro consonantal tautossilábico.

### 1.1.2 Seleção das 5 palavras menos frequentes de cada um dos corpora 1, 2 e 3

Todos os três corpora (1, 2 e 3) aqui considerados apresentam palavras com frequência de apenas uma ocorrência, que é cancelada em alguns casos e em outros não. O Corpus 1, por exemplo, apresentou 49 palavras com apenas uma ocorrência. Contudo, na escolha das 5 palavras menos frequentes dos corpora 1, 2 e 3, decidimos não trabalhar com palavras que ocorreram apenas uma única vez, já que por meio de uma só ocorrência, não é possível prever o comportamento da palavra quanto ao fenômeno em estudo.

Na verdade, como um critério geral, optamos por selecionar palavras com 4 a 6 ocorrências, visto que este intervalo é mais significativo para se prever o comportamento da palavra e, ainda, tal intervalo representa uma baixa frequência. Além disso, as palavras escolhidas com 4 a 6 ocorrências deveriam apresentar, ao mesmo tempo, 0% de cancelamento, ou seja, o fenômeno analisado não poderia se aplicar nenhuma vez nessas palavras (Cf. TAB. 1 da seção 1).

Na seleção das 5 palavras menos frequentes dos corpora 1, 2 e 3, de modo geral, excluímos nomes próprios, palavras no plural e formas verbais, especialmente no caso Corpus 1 que aborda nominais apenas. Todas essas formas são mais difíceis de ocorrer em entrevista e testes, como os que propomos aqui.

Assim, é por esse motivo que, apesar de os itens *parecer e poder* apresentarem uma baixa frequência de 4 ocorrências e, ao mesmo tempo, 0% de cancelamento – ou seja, apesar de eles se encaixarem perfeitamente no critério geral de menor frequência adotado neste trabalho –, tais itens não foram selecionados por poderem atuar como verbo também. Isso vale ainda para o item *regular* que tem 6 ocorrências e 0% de cancelamento, mas que pode ser empregado como verbo. Assim, no caso do Corpus 1, as palavras, *ar, câncer, doutor, familiar e lar* é que foram selecionadas como as 5 menos frequentes.

Vale dizer que, na seleção das 5 palavras menos frequentes dos corpora 2 e 3, tivemos de adotar o critério da maior propensão estrutural à aplicação dos fenômenos em questão, visto que, por diversas vezes, no universo de palavras observadas, havia mais do que cinco palavras para se selecionar. Desse modo, escolhemos as palavras, *cérebro, etária, silicone, diferenciação e miséria* no Corpus 2; e as palavras, *depressão, tranqüila, bronzada, complicada e traseira* no Corpus 3 (Cf. TAB. 1 da seção 1). É possível perceber que a líquida [l] está sendo focada somente uma vez em cada um destes dois últimos corpora, nas palavras *silicone* (Corpus 2) e *complicada* (Corpus 3). Isso ocorreu, porque a lateral [l] é menos frequente do que o tepe [r] nos contextos aqui considerados: posição intervocálica e encontro consonantal tautossilábico.

Como se pode notar, todas essas palavras apresentam a seqüência sonora em análise no contexto átono (pretônico ou postônico), o qual favorece mais a aplicação dos fenômenos abordados nos corpora 2 e 3 do que o contexto tônico. Inclusive, Fontes Martins (2001, p. 75) afirma: “(...) sílabas átonas ou sem proeminência podem estar sujeitas a alterações segmentais”. Não foi possível selecionar as 5 palavras menos frequentes somente com o contexto postônico que seria o preferencial, visto que, no intervalo de ocorrências observado (de 4 a 6 ocorrências), não havia muitas opções de palavras para os corpora 2 e 3.

Das 5 palavras menos frequentes selecionadas no Corpus 2, três palavras apresentam a líquida intervocálica em contexto postônico, que é o maior favorecedor à aplicação do fenômeno do cancelamento da líquida intervocálica. As outras duas palavras têm a líquida intervocálica em contexto pretônico, que também favorece mais o cancelamento do que o contexto tônico (Cf. FONTES MARTINS, 2001).

Já no Corpus 3, todas as 5 palavras menos frequentes apresentam a líquida em encontro consonantal tautossilábico na posição pretônica. No caso do fenômeno do Corpus 3 também o contexto postônico seria o maior favorecedor, contudo, o contexto pretônico, como um ambiente átono (Cf. CRISTÓFARO SILVA, 2003b), ainda se mostra relevante quanto à aplicação do fenômeno. Por esse motivo, pode-se falar que o critério da maior propensão foi adotado no caso do Corpus 3 e, ainda, no caso do Corpus 2.

Todas essas opções metodológicas apresentadas até o momento visaram fornecer a esta pesquisa uma seleção apropriada das palavras mais e menos frequentes dos corpora 1, 2 e 3, para a elaboração da entrevista e dos testes a serem aplicados na coleta de dados. Considerou-se importante neste trabalho fazer também uma pesquisa da frequência dessas palavras selecionadas em corpora eletrônicos do português brasileiro. Isso é feito na próxima seção com o intuito de avaliar o comportamento das palavras selecionadas em corpora que são mais representativos da língua por apresentarem um número bastante significativo de palavras.

## 1.2. Avaliação, em corpora eletrônicos, dos dados obtidos dos corpora 1, 2 e 3

Um outro procedimento metodológico adotado neste estudo foi analisar a frequência dessas 30 palavras da tabela 1 (Cf. seção 1) nos corpora eletrônicos do CETEM/ Público (Corpus de Extratos de Textos Eletrônicos MCT/ Público)

– mais especificamente, o Corpus NILC/ Universidade de São Carlos – e do LAEL (Linguística Aplicada e Estudos da Linguagem) – escrita e fala<sup>59</sup>. Estes corpora foram escolhidos, por terem sido, na época em que este estudo foi realizado, os melhores (em termos de acessibilidade e quantidade de dados) corpora eletrônicos disponíveis.

Para conferir os resultados das pesquisas de frequência nos corpora do LAEL-escrita e fala e do NILC/ São Carlos, para 30 das 50 palavras utilizadas neste trabalho (Cf. TAB. 1 da seção 1), apresentamos a tabela 2 que também mostra a frequência destas 30 palavras nos corpora 1, 2 ou 3, analisados neste estudo. É preciso dizer que tais palavras se apresentam na tabela, em ordem decrescente de frequência quanto aos corpora 1, 2 e 3.

TABELA 2

Frequência das 30 palavras no Corpus analisado (corpora 1, 2 ou 3) e nos corpora do LAEL-escrita e fala, do NILC/ São Carlos.

Corpora	Grau de frequência	Palavras	Frequência no Corpus analisado (corpora 1, 2 ou 3)	Frequência no Corpus do NILC/ São Carlos	Frequência no Corpus do LAEL-escrita	Frequência no Corpus do LAEL-fala
CORPUS 1: Huback (2003)	Mais frequentes	1 Melhor	93	14.530	17.145	1.466
		2 Lugar	89	10.200	10.808	1.469
		3 Mulher	74	7.975	11.074	1.018
		4 Qualquer	40	13.725	13.486	1.058
		5 Computador	35	2.926	3.042	99
	Menos frequentes	6 Ar	6	4.011	5.333	100
		7 Câncer	4	1.004	1.576	73
		8 Doutor	4	700	1.218	324
		9 Familiar	4	900	1.002	45
		10 Lar	4	179	347	68
SUBTOTAL			10 palavras			
CORPUS 2: Fontes Martins (2001)	Mais frequentes	11 Para	859	297.628	353.847	3.611
		12 Ele	214	45.190	71.290	23.104
		13 Ela	169	16.903	25.361	13.126
		14 Eles	144	12.767	20.136	14.145
		15 Óculos	40	441	627	22
	Menos frequentes	16 Cérebro	5	777	732	26
		17 Etária	5	170	192	28
		18 Silicone	5	170	227	0
		19 Diferenciação	4	152	97	1
		20 Miséria	4	975	1.437	66
SUBTOTAL			10 palavras			
CORPUS 3: Cristóvão Silva (2003a)	Mais frequentes	21 Problema	77	8.920	9.032	1.631
		22 Outro	58	14.926	17.686	3.549
		23 Quatro	57	15.028	18.337	2.103
		24 Sempre	51	13.667	15.513	5.690
		25 Primeiro	48	20.011	23.822	1.641
	Menos frequentes	26 Depressão	5	291	430	32
		27 Tranquila	5	79	63	77
		28 Bronzeada	4	17	25	0
		29 Complicada	4	314	325	24
		30 Traseira	4	280	431	2
SUBTOTAL			10 palavras			
TOTAL CERPA			30 palavras			

Fonte: <http://www.linguateca.pt/CETEMPublico/> e <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-30 mar. 2004) e Huback (2003), Fontes Martins (2001) e Cristóvão Silva (2003a)

Esta pesquisa de frequência nos corpora do LAEL-escrita e fala e do NILC/ São Carlos, para as 30 das 50 palavras utilizadas neste trabalho, é interessante por fornecer uma visão mais ampla da frequência dessas palavras na língua. É sabido que uma palavra pode ser muito frequente em um corpus particular, mas não ter grande frequência na língua. Certamente que corpora substanciais, como o LAEL-escrita e fala e o NILC/ São Carlos, podem oferecer parâmetros de frequência mais próximos da realidade da língua.

Esse é o motivo maior pelo qual propusemos a análise feita na tabela 2 que repete, da primeira até a quarta coluna, dados da tabela 1. Contudo, a tabela 2 apresenta ainda, como se disse, os valores de frequência das 30 palavras dos corpora 1, 2 e 3, nos corpora do LAEL-escrita e fala e do NILC/ São Carlos. Isso é feito da quinta à sétima coluna da tabela 2.

Como se pode ver nesta tabela, os resultados de frequência das 30 palavras dos corpora 1, 2 e 3 em relação aos corpora eletrônicos do LAEL-escrita e fala e do NILC/ São Carlos são diferenciados. Há casos em que os valores se mostram semelhantes em todos os corpora (corpora 1, 2 e 3, corpus LAEL-escrita e fala e corpus NILC/ São Carlos). Isso ocorre, por exemplo, com as palavras de número 1 e 2 (Cf. TAB. 2), em que todos os corpora apresentam valores mais altos para essas palavras.

<sup>59</sup> Os endereços eletrônicos do CETEM/Público e do LAEL (escrita e fala) são <<http://www.linguateca.pt/CETEMPublico/>> e <<http://LAEL.pucsp.br/corpora/index.htm>>, respectivamente. Maiores informações sobre estes dois corpora podem ser obtidas nesses endereços eletrônicos.

Ao mesmo tempo, há também casos em que uma única palavra apresenta um valor de frequência alto em um corpus e em outro corpus não. Este é o caso da palavra de número 5 (Cf. TAB. 2), *computador*, que apresenta um valor de frequência mais para alto no Corpus 1 em análise e valores de frequência relativamente mais baixos nos corpora do LAEL-escrita e fala e do NILC/ São Carlos.

É possível, no entanto, que esses baixos valores para a palavra *computador* se devam ao fato de que tal palavra é relativamente recente na língua, e os corpora eletrônicos aqui considerados abarcam também dados mais antigos. A palavra em questão, *computador*, passou a ser mais empregada na nossa língua, a partir da década de oitenta, com o início da popularização do computador no Brasil. Contudo, pode-se afirmar pela avaliação da tabela 5 que, de modo geral, os valores de frequência (altos ou baixos) encontrados nos corpora 1, 2 e 3 são confirmados pelos corpora eletrônicos do LAEL-escrita e fala e do NILC/ São Carlos.

### 1.3 Apresentação dos dados obtidos do LAEL-escrita e fala

Adicionalmente à informação dos valores de frequência – nos corpora 1, 2 e 3 e nos corpora do LAEL-escrita e fala e do NILC/ São Carlos – das 30 palavras apresentadas na tabela 2, também selecionamos no LAEL-escrita e fala, as 5 palavras mais frequentes e as 5 palavras menos frequentes para cada um dos 3 fenômenos contemplados pelos corpora 1, 2 e 3. Na verdade, buscou-se selecionar as 5 primeiras palavras de cada fenômeno que fossem mais frequentes no LAEL-escrita e no LAEL-fala, concomitantemente. De modo semelhante, isso foi feito em relação às 5 palavras menos frequentes de cada fenômeno.

Assim, a nova análise contou inicialmente com um total de 30 palavras: 15 palavras mais frequentes e 15 palavras menos frequentes; ou ainda, dito de outra forma, 10 palavras com “r” final em nominais, 10 palavras com líquida intervocálica e 10 palavras com encontro consonantal tautossilábico.

A seleção dessas novas 30 palavras não pôde ser feita para o Corpus do NILC/São Carlos, visto que neste corpus não é possível ver as frequências de todas as suas palavras, uma a uma, o que é possível no LAEL-escrita e fala. No corpus do NILC/ São Carlos, só é possível fazer consulta da frequência de palavras pré-selecionadas, indicadas pelo usuário desse corpus no *site*. Essas 30 novas palavras mais e menos frequentes do corpus do LAEL-escrita e fala são apresentadas na tabela 3 abaixo, na ordem decrescente de frequência em que ocorreram no LAEL-escrita:

TABELA 3

As 30 palavras mais e menos freqüentes do LAEL-escrita e fala, distribuídas de acordo com os fenômenos dos corpora 1, 2 e 3

Fenômenos	Grau de Freqüência	Palavras	Valores de freqüência no LAEL-escrita	Valores de freqüência no LAEL-fala
Cancelamento de “r” final em nominais (Corpus 1)	Mais freqüentes	1. Maior	24.534	896
		2. Melhor	17.145	1.466
		3. Qualquer	13.486	1.058
		4. Mulher	11.074	1.018
		5. Lugar	10.808	1.469
	Menos freqüentes	6. Impostor	14	1
		7. Incentivador	13	1
		8. Instalador	13	1
		9. Alternador	11	1
		10. Falsificador	11	1
SUBTOTAL		10 palavras		
Cancelamento de líquida intervocálica (Corpus 2)	Mais freqüentes	11. Ele	71.290	23.104
		12. Era	32.247	18.847
		13. Ela	25.361	13.126
		14. Eles	20.136	14.145
		15. Agora	19.373	8.145
	Menos freqüentes	16. Capivara	15	3
		17. Filmadora	15	4
		18. Cítara	13	7
		19. Versículo	13	6
		20. Marmelo	11	10
SUBTOTAL		10 palavras		
Cancelamento de encontro consonantal tautossilábico (Corpus 3)	Mais freqüentes	21. Grande	24.914	2.553
		22. Primeiro	23.822	1.641
		23. Outros	20.334	1.681
		24. Quatro	18.337	2.103
		25. Outro	17.686	3.549
	Menos freqüentes	26. Apodrecer	15	5
		27. Platônico	15	4
		28. Sobreaviso	15	3
		29. Braçal	12	7
		30. Supérflua	12	5
SUBTOTAL		10 palavras		
TOTAL GERAL		30 palavras		

Fonte: <http://www.linguateca.pt/CETEMPublico/> e <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-26 de fevereiro de 2004) e Huback (2003), Fontes Martins (2001) e Cristófaros Silva (2003a)

A tabela 3 apresenta, na terceira coluna, as 30 palavras mais freqüentes (15 palavras) e menos freqüentes (15 palavras) retiradas do LAEL-escrita e fala. A seleção dessas palavras considerou os fenômenos focalizados pelos corpora 1, 2 e 3, o que pode ser visto na primeira coluna. A segunda coluna discrimina o grau de freqüência (mais ou menos freqüente) dessas palavras. A quarta e a quinta colunas apresentam os valores de freqüência das 30 palavras no LAEL-escrita e do LAEL-fala, respectivamente.

Vale dizer que a seleção dessas outras 30 palavras foi relativamente simples. O *site* do LAEL disponibiliza esse utilitário com fácil acessibilidade. Especialmente no caso das palavras mais freqüentes, a escolha foi bastante exata, sem a necessidade de se adotar nenhum critério para nortear a seleção. Bastou observar quais eram as 5 primeiras palavras mais freqüentes ao mesmo tempo, no LAEL-escrita e no LAEL-fala. Maiores esclarecimentos sobre a seleção das palavras mais freqüentes são dados a seguir, na seção 1.3.1. Já no caso das palavras menos freqüentes, tivemos de adotar alguns critérios adicionais que são descritos na seção 1.3.2.

### 1.3.1 Especificidades na seleção das palavras mais freqüentes do LAEL-escrita e fala

Como se pode perceber pela análise das 15 palavras mais freqüentes do LAEL-escrita e fala na tabela 6 da seção anterior, há 10 palavras (Cf. palavras de número 2, 3, 4, 5, 11, 13, 14, 22, 24, 25 da TAB. 3) que coincidem com palavras mais freqüentes dos corpora 1, 2 e 3 (Cf. palavras de número 1, 2, 3, 4, 12, 13, 14, 22, 23, 25 da TAB. 1 na seção 1). Assim, ao se comparar as tabelas 1 e 3, nota-se que:

- quanto ao fenômeno de cancelamento de “r” final (Corpus 1), quatro palavras da tabela 3 coincidem com palavras da tabela 1: *melhor*, *qualquer*, *mulher* e *lugar*;
- quanto ao fenômeno de cancelamento de líquidas intervocálicas (Corpus 2), três palavras da tabela 3 coincidem com palavras da tabela 1: *ele*, *ela* e *eles*;

- quanto ao fenômeno de cancelamento líquida em encontro consonantal tautossilábico (Corpus 3), três palavras da tabela 3 coincidem com palavras da tabela 1: *primeiro*, *quatro* e *outro*.

Essa repetição na tabela 6 só ocorre com as palavras mais freqüentes. Tendo em vista o fato dessa repetição de 10 palavras, a presente pesquisa contará com 50 palavras: as 30 que foram apresentadas na tabela 1 (Cf. seção 1) e as 20 palavras da tabela 3 (seção 1.3) que não repetem as palavras da tabela 1. Destas 20 palavras, somente 5 são mais freqüentes: *maior*, *era*, *agora*, *grande* e *outros*. As outras 15 palavras menos freqüentes são tratadas na próxima seção.

### 1.3.2 Especificidades na seleção das palavras menos freqüentes do LAEL-escrita e fala

Como se pode notar pela análise das palavras menos freqüentes na tabela 3 (Cf. seção 1.3), selecionamos palavras com, simultaneamente, de 11 a 15 ocorrências no LAEL-escrita, e de 1 a 10 ocorrências no LAEL-fala. Esse foi um primeiro critério mais geral. O LAEL-escrita é mais abundante em dados do que o LAEL-fala. É por esse motivo que o intervalo de ocorrências para o LAEL-escrita é superior ao intervalo de ocorrências para o LAEL-fala.

O segundo critério adotado foi o de a palavra apresentar ambiente estrutural favorável à aplicação do fenômeno em questão. Ainda, no caso dos fenômenos dos corpora 2 e 3, buscou-se selecionar, pelo menos 2 palavras com a líquida [l] que foi mais difícil de ser encontrada, sendo bem menos freqüente do que a líquida [r]. Dessa forma, para cada um dos fenômenos dos corpora 2 e 3, conseguimos selecionar 2 palavras com a líquida [l] e 3 palavras com a líquida [r], no LAEL-escrita e fala.

Na seleção das palavras menos freqüentes da tabela 3 (Cf. seção 1.3), de forma geral, evitamos selecionar: nomes próprios, gentílicos, estrangeirismos mais marcados, termos caracterizadamente técnicos e palavras chulas. Excluímos tais formas, porque elas são mais difíceis de ocorrer em nossos testes e entrevista que serão formulados com base nas 30 palavras da tabela 1 (Cf. seção 1) e nas 20 da tabela 3 (Cf. seção 1.3) que não repetem dados da tabela 1.

Procuramos adotar a forma como os dicionários fazem a notação das entradas de suas palavras, por isso evitamos formas verbais flexionadas, nomes plurais e/ ou femininos, os quais também seriam mais difíceis de ocorrer nos nossos testes e entrevista.

Evitamos palavras com mais de um ambiente estrutural para mais de um fenômeno observado (corpora 1, 2 e 3). Este seria, por exemplo, o caso da palavra *trabalhador*, que apresenta, ao mesmo tempo, ambiente para a aplicação do fenômeno do Corpus 3 (cancelamento de líquidas em encontro consonantal tautossilábico) e do fenômeno do Corpus 1 (cancelamento de “r” final em nominais). Contudo, deve-se dizer que, ainda assim, não foi possível evitar essa situação por completo, no caso das palavras menos freqüentes do LAEL-escrita e fala. Há um único caso, a palavra *instalador* (Cf. TAB. 3, seção 1.3) que apresenta, respectivamente, um ambiente estrutural (-*ala*-) para o fenômeno do Corpus 2 (cancelamento de líquidas intervocálicas) e um ambiente estrutural (-*dor*) para o fenômeno do Corpus 1 (cancelamento de “r” final em nominais).

Ainda, evitamos palavras com mais de um ambiente estrutural para o mesmo fenômeno. Como exemplo, podemos citar a palavra *alarido* que apresenta dois ambientes estruturais para a aplicação do fenômeno do Corpus 2 (cancelamento de líquidas intervocálicas). Essas duas últimas medidas foram tomadas pelo fato de que seria mais difícil controlar, na pesquisa, palavras deste tipo, com duas estruturas, ao mesmo tempo, passíveis de análise.

O motivo de termos selecionado, no LAEL-escrita e fala, as palavras menos freqüentes dos três fenômenos em análise, com base no intervalo de 11 a 15 ocorrências no LAEL-escrita e de 1 a 10 ocorrências no LAEL-fala, é justamente pelo fato de ter havido muitos critérios de seleção e exclusão de palavras, os quais foram explicitados logo acima. Com todos esses critérios, foi preciso alargar o número de ocorrências, especialmente no LAEL-fala, para se conseguir escolher as 5 palavras menos freqüentes em relação a cada um dos 3 fenômenos aqui analisados. Especialmente em relação ao fenômeno do Corpus 1 que apresenta um contexto estrutural menos freqüente.

Assim, quanto ao fenômeno do Corpus 1 (cancelamento de “r” final em nominais), selecionamos, no LAEL-escrita e fala, as palavras: *impostor*, *incentivador*, *instalador*, *alternador* e *falsificador* (Cf. TAB. 3, seção 1.3). Note-se que todas essas palavras apresentam ambiente favorável à aplicação do fenômeno, ou seja, apresentam ou o sufixo -*dor*, ou o sufixo -*tor* (Cf. HUBACK, 2003). Note-se ainda, pela análise dessas palavras na tabela 3, que somente elas apresentam no LAEL-fala apenas 1 ocorrência. Como foi destacado no parágrafo anterior, isso se deve ao fato de que nominais com “r” final são menos freqüentes, especialmente, em relação às palavras dos fenômenos do Corpus 2 e 3 que, ao contrário, são mais freqüentes.

Quanto ao fenômeno do Corpus 2 (cancelamento de líquidas intervocálicas), as palavras selecionadas foram: *capivara*, *filmadora*, *citara*, *versículo* e *marmelo* (Cf. TAB. 3, seção 1.3). Note-se que todas essas palavras apresentam ambiente estrutural favorável à aplicação do fenômeno em questão: a posição postônica.

Em relação ao fenômeno do Corpus 3 (cancelamento de líquidas em encontro consonantal tautossilábico), as palavras escolhidas foram: *apodrecer*, *platônico*, *sobreaviso*, *braçal* e *supérflua*. No caso deste fenômeno, todas as palavras

apresentam o ambiente átono favorecedor. Contudo, como não foi possível encontrar, adotando-se todos os critérios explicitados acima, 5 palavras com o ambiente átono preferencial, o contexto postônico, tivemos de selecionar, 4 dessas 5 palavras com o ambiente pretônico. De toda forma, o ambiente pretônico ainda é um favorecedor à aplicação do fenômeno do cancelamento de líquida em encontro consonantal tautossilábico (Cf. CRISTÓFARO SILVA, 2003b).

#### 1.4 Apresentação geral das palavras consideradas nesta pesquisa

Este estudo contará com 50 palavras no total, as quais são apresentadas na tabela 4 abaixo:

**TABELA 4**  
As 50 palavras consideradas nesta pesquisa

Corpora	Grau de	Palavra	Fenômeno
<b>CORPUS 1: Huback (2003)</b>	+ Mais freqüentes	1. Computador	<b>Cancelamento de “r” em final de palavra</b>
		2. Lugar	
		3. Melhor	
		4. Mulher	
		5. Qualquer	
	- Menos freqüentes	6. Ar	
		7. Câncer	
		8. Doutor	
		9. Familiar	
		10. Lar	
<b>CORPUS 2: Fontes Martins (2001)</b>	+ Mais freqüentes	11. Ela	<b>Cancelamento de líquida intervocálica</b>
		12. Ele	
		13. Eles	
		14. Óculos	
		15. Para	
	- Menos freqüentes	16. Cérebro	
		17. Diferenciação	
		18. Etária	
		19. Miséria	
		20. Silicone	
<b>CORPUS 3: Cristóforo Silva (2003a)</b>	+ Mais freqüentes	21. Outro	<b>Cancelamento de encontro consonantal tautossilábico</b>
		22. Primeiro	
		23. Problema	
		24. Quatro	
		25. Sempre	
	- Menos freqüentes	26. Bronzeada	
		27. Complicada	
		28. Depressão	
		29. Tranqüila	
		30. Traseira	
<b>LAEL-escrita e fala</b>	+ Mais freqüentes	31. Maior	<b>Cancelamento de “r” em final de palavra</b>
		32. Agora	<b>Cancelamento de líquida intervocálica</b>
		33. Era	<b>Cancelamento de encontro consonantal tautossilábico</b>
		34. Grande	
		35. Outros	
	- Menos freqüentes	36. Alternador	<b>Cancelamento de “r” em final de palavra</b>
		37. Falsificador	
		38. Impostor	
		39. Incentivador	
		40. Instalador	
		41. Capivara	<b>Cancelamento de líquida intervocálica</b>
		42. Cítara	
		43. Filmadora	
		44. Marmelo	
		45. Versículo	
		46. Apodrecer	<b>Cancelamento de encontro consonantal tautossilábico</b>
		47. Braçal	
		48. Platônico	
		49. Sobreaviso	
		50. Supérflua	

Fonte: <http://www.linguateca.pt/CETEMPublico/> e <http://LAEL.pucsp.br/corpora/index.htm> (Acesso: 17-26 de fevereiro de 2004) e Huback (2003), Fontes Martins (2001) e Cristóforo Silva (2003a)

A tabela 4 acima apresenta, na terceira coluna, as 50 palavras a serem consideradas no presente trabalho. A primeira coluna mostra o corpus de onde cada palavra foi selecionada, e a segunda coluna demonstra o grau de freqüência (mais ou menos freqüente) de cada uma das 50 palavras. A última coluna discrimina o fenômeno a que cada palavra está relacionada.

Como se pode perceber pela análise da tabela 4, há uma certa assimetria nos dados do LAEL-escrita e fala (Cf. palavras do número 36 ao 50 da TAB. 4), em relação à frequência. Enquanto para cada um dos corpora 1, 2 e 3, há exatamente 5 palavras mais frequentes e 5 palavras menos frequentes, para o LAEL-escrita e fala, há 5 palavras mais frequentes e 15 palavras menos frequentes. Vale lembrar que essa assimetria nos dados desses corpora eletrônicos se deve ao fato de que tivemos de excluir 10 palavras (Cf. palavras de número 2, 3, 4, 5, 11, 13, 14, 22, 24, 25 da TAB. 3) que coincidiam com palavras mais frequentes dos corpora 1, 2 e 3 (Cf. palavras de número 1, 2, 3, 4, 12, 13, 14, 22, 23, 25 da TAB. 1 na seção 1). Essa repetição ocorreu somente com as palavras mais frequentes. Maiores detalhes podem ser conferidos na seção 1.3.1, apresentada anteriormente.

## APÊNDICE F

### Tabelas – palavras e indivíduos do estudo transversal

Abaixo, apresentamos tabelas com os percentuais de cancelamento de cada um dos 12 indivíduos do estudo transversal quanto às 30 palavras avaliadas em tal estudo.

#### Indivíduo 1 – L1W

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	6	54,5%	5	45,5%	11
lugar	3	23,1%	10	76,9%	13
mulher	5	29,4%	12	70,6%	17
qualquer	0	,0%	4	100,0%	4
computador	3	75,0%	1	25,0%	4
maior	3	21,4%	11	78,6%	14
mar	5	62,5%	3	37,5%	8
governador	2	50,0%	2	50,0%	4
jogador	3	50,0%	3	50,0%	6
dólar	0	,0%	6	100,0%	6
para	58	51,8%	54	48,2%	112
ele	46	64,8%	25	35,2%	71
ela	20	83,3%	4	16,7%	24
eles	5	14,7%	29	85,3%	34
óculos	4	50,0%	4	50,0%	8
era	21	87,5%	3	12,5%	24
dinheiro	8	72,7%	3	27,3%	11
feira	7	100,0%	0	,0%	7
agora	7	70,0%	3	30,0%	10
Lula	10	100,0%	0	,0%	10
problema	0	,0%	6	100,0%	6
outro	9	47,4%	10	52,6%	19
quatro	4	100,0%	0	,0%	4
sempre	6	50,0%	6	50,0%	12
primeiro	8	88,9%	1	11,1%	9
Brasil	5	100,0%	0	,0%	5
presidente	6	100,0%	0	,0%	6
três	8	100,0%	0	,0%	8
grande	8	100,0%	0	,0%	8
livro	5	71,4%	2	28,6%	7
TOTAL	275	57,1%	207	42,9%	482

#### Indivíduo 2 – L1V

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	8	100,0%	0	,0%	8
lugar	8	80,0%	2	20,0%	10
mulher	6	75,0%	2	25,0%	8
qualquer	1	25,0%	3	75,0%	4
computador	5	100,0%	0	,0%	5
maior	6	85,7%	1	14,3%	7
mar	4	100,0%	0	,0%	4
governador	4	100,0%	0	,0%	4
jogador	4	66,7%	2	33,3%	6
dólar	6	100,0%	0	,0%	6
para	43	69,4%	19	30,6%	62
ele	43	89,6%	5	10,4%	48
ela	25	89,3%	3	10,7%	28
eles	12	70,6%	5	29,4%	17
óculos	5	62,5%	3	37,5%	8
era	23	95,8%	1	4,2%	24
dinheiro	7	77,8%	2	22,2%	9
feira	4	100,0%	0	,0%	4
agora	12	80,0%	3	20,0%	15
Lula	8	100,0%	0	,0%	8
problema	1	7,1%	13	92,9%	14
outro	1	12,5%	7	87,5%	8
quatro	1	25,0%	3	75,0%	4
sempre	9	69,2%	4	30,8%	13
primeiro	7	77,8%	2	22,2%	9
Brasil	10	100,0%	0	,0%	10
presidente	5	100,0%	0	,0%	5
três	8	88,9%	1	11,1%	9
grande	4	100,0%	0	,0%	4
livro	4	40,0%	6	60,0%	10
TOTAL	284	76,5%	87	23,5%	371

## Indivíduo 3 – N1A

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	9	81,8%	2	18,2%	11
lugar	7	100,0%	0	,0%	7
mulher	7	77,8%	2	22,2%	9
qualquer	6	85,7%	1	14,3%	7
computador	5	100,0%	0	,0%	5
maior	4	66,7%	2	33,3%	6
mar	4	80,0%	1	20,0%	5
governador	4	100,0%	0	,0%	4
jogador	7	77,8%	2	22,2%	9
dólar	4	100,0%	0	,0%	4
para	50	96,2%	2	3,8%	52
ele	43	95,6%	2	4,4%	45
ela	14	100,0%	0	,0%	14
eles	10	100,0%	0	,0%	10
óculos	6	100,0%	0	,0%	6
era	5	100,0%	0	,0%	5
dinheiro	12	92,3%	1	7,7%	13
feira	9	100,0%	0	,0%	9
agora	13	100,0%	0	,0%	13
Lula	4	100,0%	0	,0%	4
problema	13	100,0%	0	,0%	13
outro	10	76,9%	3	23,1%	13
quatro	2	50,0%	2	50,0%	4
sempre	7	100,0%	0	,0%	7
primeiro	5	100,0%	0	,0%	5
Brasil	9	100,0%	0	,0%	9
presidente	4	100,0%	0	,0%	4
três	6	100,0%	0	,0%	6
grande	9	100,0%	0	,0%	9
livro	5	83,3%	1	16,7%	6
TOTAL	293	93,3%	21	6,7%	314

## Indivíduo 4 –

N1L PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	9	100,0%	0	,0%	9
lugar	12	100,0%	0	,0%	12
mulher	7	63,6%	4	36,4%	11
qualquer	12	85,7%	2	14,3%	14
computador	8	80,0%	2	20,0%	10
maior	6	85,7%	1	14,3%	7
mar	12	92,3%	1	7,7%	13
governador	5	100,0%	0	,0%	5
jogador	3	60,0%	2	40,0%	5
dólar	4	100,0%	0	,0%	4
para	99	92,5%	8	7,5%	107
ele	113	97,4%	3	2,6%	116
ela	58	100,0%	0	,0%	58
eles	13	86,7%	2	13,3%	15
óculos	5	100,0%	0	,0%	5
era	40	100,0%	0	,0%	40
dinheiro	16	100,0%	0	,0%	16
feira	15	100,0%	0	,0%	15
agora	34	94,4%	2	5,6%	36
Lula	5	100,0%	0	,0%	5
problema	21	100,0%	0	,0%	21
outro	17	73,9%	6	26,1%	23
quatro	6	100,0%	0	,0%	6
sempre	10	66,7%	5	33,3%	15
primeiro	8	100,0%	0	,0%	8
Brasil	7	100,0%	0	,0%	7
presidente	7	100,0%	0	,0%	7
três	18	100,0%	0	,0%	18
grande	6	100,0%	0	,0%	6
livro	11	91,7%	1	8,3%	12
TOTAL	587	93,8%	39	6,2%	626

## Indivíduo 5 – L2L

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	9	69,2%	4	30,8%	13
lugar	5	100,0%	0	,0%	5
mulher	6	60,0%	4	40,0%	10
qualquer	3	60,0%	2	40,0%	5
computador	3	60,0%	2	40,0%	5
maior	6	100,0%	0	,0%	6
mar	7	87,5%	1	12,5%	8
governador	3	60,0%	2	40,0%	5
jogador	4	57,1%	3	42,9%	7
dólar	8	88,9%	1	11,1%	9
para	25	43,1%	33	56,9%	58
ele	23	23,7%	74	76,3%	97
ela	15	93,8%	1	6,3%	16
eles	8	72,7%	3	27,3%	11
óculos	5	71,4%	2	28,6%	7
era	25	96,2%	1	3,8%	26
dinheiro	10	66,7%	5	33,3%	15
feira	6	100,0%	0	,0%	6
agora	16	64,0%	9	36,0%	25
Lula	6	100,0%	0	,0%	6
problema	11	100,0%	0	,0%	11
outro	4	50,0%	4	50,0%	8
quatro	4	100,0%	0	,0%	4
sempre	2	50,0%	2	50,0%	4
primeiro	7	87,5%	1	12,5%	8
Brasil	5	100,0%	0	,0%	5
presidente	8	100,0%	0	,0%	8
três	5	100,0%	0	,0%	5
grande	5	83,3%	1	16,7%	6
livro	5	71,4%	2	28,6%	7
TOTAL	249	61,3%	157	38,7%	406

## Indivíduo 6 – L2R

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	11	78,6%	3	21,4%	14
lugar	4	80,0%	1	20,0%	5
mulher	6	100,0%	0	,0%	6
qualquer	5	71,4%	2	28,6%	7
computador	4	80,0%	1	20,0%	5
maior	4	100,0%	0	,0%	4
mar	6	100,0%	0	,0%	6
governador	4	80,0%	1	20,0%	5
jogador	3	75,0%	1	25,0%	4
dólar	4	80,0%	1	20,0%	5
para	17	70,8%	7	29,2%	24
ele	17	60,7%	11	39,3%	28
ela	4	100,0%	0	,0%	4
eles	3	75,0%	1	25,0%	4
óculos	7	100,0%	0	,0%	7
era	4	100,0%	0	,0%	4
dinheiro	7	87,5%	1	12,5%	8
feira	5	83,3%	1	16,7%	6
agora	9	60,0%	6	40,0%	15
Lula	5	100,0%	0	,0%	5
problema	7	100,0%	0	,0%	7
outro	1	25,0%	3	75,0%	4
quatro	4	100,0%	0	,0%	4
sempre	6	85,7%	1	14,3%	7
primeiro	7	87,5%	1	12,5%	8
Brasil	7	87,5%	1	12,5%	8
presidente	6	100,0%	0	,0%	6
três	3	75,0%	1	25,0%	4
grande	6	100,0%	0	,0%	6
livro	4	80,0%	1	20,0%	5
TOTAL	180	80,0%	45	20,0%	225

## Indivíduo 7 – N2L

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	9	100,0%	0	,0%	9
lugar	4	100,0%	0	,0%	4
mulher	6	100,0%	0	,0%	6
qualquer	10	100,0%	0	,0%	10
computador	6	100,0%	0	,0%	6
maior	4	100,0%	0	,0%	4
mar	21	100,0%	0	,0%	21
governador	5	100,0%	0	,0%	5
jogador	4	80,0%	1	20,0%	5
dólar	10	90,9%	1	9,1%	11
para	65	98,5%	1	1,5%	66
ele	83	100,0%	0	,0%	83
ela	22	100,0%	0	,0%	22
eles	7	87,5%	1	12,5%	8
óculos	8	100,0%	0	,0%	8
era	13	100,0%	0	,0%	13
dinheiro	22	100,0%	0	,0%	22
feira	9	100,0%	0	,0%	9
agora	18	100,0%	0	,0%	18
Lula	14	100,0%	0	,0%	14
problema	8	100,0%	0	,0%	8
outro	6	85,7%	1	14,3%	7
quatro	4	80,0%	1	20,0%	5
sempre	6	85,7%	1	14,3%	7
primeiro	7	100,0%	0	,0%	7
Brasil	7	100,0%	0	,0%	7
presidente	5	100,0%	0	,0%	5
três	12	100,0%	0	,0%	12
grande	8	100,0%	0	,0%	8
livro	5	100,0%	0	,0%	5
TOTAL	408	98,3%	7	1,7%	415

## Indivíduo 8 – N2M

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	12	100,0%	0	,0%	12
lugar	5	100,0%	0	,0%	5
mulher	4	100,0%	0	,0%	4
qualquer	7	100,0%	0	,0%	7
computador	7	87,5%	1	12,5%	8
maior	4	100,0%	0	,0%	4
mar	10	100,0%	0	,0%	10
governador	5	100,0%	0	,0%	5
jogador	4	100,0%	0	,0%	4
dólar	4	100,0%	0	,0%	4
para	78	96,3%	3	3,7%	81
ele	56	100,0%	0	,0%	56
ela	29	100,0%	0	,0%	29
eles	32	100,0%	0	,0%	32
óculos	4	80,0%	1	20,0%	5
era	14	93,3%	1	6,7%	15
dinheiro	5	100,0%	0	,0%	5
feira	8	100,0%	0	,0%	8
agora	10	90,9%	1	9,1%	11
Lula	11	100,0%	0	,0%	11
problema	8	100,0%	0	,0%	8
outro	7	100,0%	0	,0%	7
quatro	6	100,0%	0	,0%	6
sempre	5	100,0%	0	,0%	5
primeiro	7	100,0%	0	,0%	7
Brasil	11	100,0%	0	,0%	11
presidente	4	100,0%	0	,0%	4
três	12	100,0%	0	,0%	12
grande	4	100,0%	0	,0%	4
livro	6	100,0%	0	,0%	6
TOTAL	379	98,2%	7	1,8%	386

## Indivíduo 9 – L3R

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	14	93,3%	1	6,7%	15
lugar	17	100,0%	0	,0%	17
mulher	24	68,6%	11	31,4%	35
qualquer	10	66,7%	5	33,3%	15
computador	9	100,0%	0	,0%	9
maior	7	87,5%	1	12,5%	8
mar	5	100,0%	0	,0%	5
governador	4	100,0%	0	,0%	4
jogador	4	100,0%	0	,0%	4
dólar	8	100,0%	0	,0%	8
para	120	63,5%	69	36,5%	189
ele	124	87,3%	18	12,7%	142
ela	47	87,0%	7	13,0%	54
eles	27	69,2%	12	30,8%	39
óculos	4	100,0%	0	,0%	4
era	16	94,1%	1	5,9%	17
dinheiro	24	92,3%	2	7,7%	26
feira	4	100,0%	0	,0%	4
agora	34	89,5%	4	10,5%	38
Lula	7	100,0%	0	,0%	7
problema	15	93,8%	1	6,3%	16
outro	7	70,0%	3	30,0%	10
quatro	5	45,5%	6	54,5%	11
sempre	11	68,8%	5	31,3%	16
primeiro	4	100,0%	0	,0%	4
Brasil	4	100,0%	0	,0%	4
presidente	13	92,9%	1	7,1%	14
três	15	100,0%	0	,0%	15
grande	13	100,0%	0	,0%	13
livro	3	75,0%	1	25,0%	4
TOTAL	599	80,2%	148	19,8%	747

## Indivíduo 10 – L3J

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	5	83,3%	1	16,7%	6
lugar	4	40,0%	6	60,0%	10
mulher	8	57,1%	6	42,9%	14
qualquer	5	83,3%	1	16,7%	6
computador	4	50,0%	4	50,0%	8
maior	4	66,7%	2	33,3%	6
mar	3	60,0%	2	40,0%	5
governador	3	50,0%	3	50,0%	6
jogador	3	75,0%	1	25,0%	4
dólar	2	33,3%	4	66,7%	6
para	43	62,3%	26	37,7%	69
ele	20	71,4%	8	28,6%	28
ela	13	72,2%	5	27,8%	18
eles	14	46,7%	16	53,3%	30
óculos	2	40,0%	3	60,0%	5
era	10	90,9%	1	9,1%	11
dinheiro	12	92,3%	1	7,7%	13
feira	5	100,0%	0	,0%	5
agora	9	100,0%	0	,0%	9
Lula	4	100,0%	0	,0%	4
problema	0	,0%	7	100,0%	7
outro	3	33,3%	6	66,7%	9
quatro	2	50,0%	2	50,0%	4
sempre	5	41,7%	7	58,3%	12
primeiro	6	100,0%	0	,0%	6
Brasil	4	100,0%	0	,0%	4
presidente	4	100,0%	0	,0%	4
três	8	100,0%	0	,0%	8
grande	5	100,0%	0	,0%	5
livro	3	60,0%	2	40,0%	5
TOTAL	213	65,1%	114	34,9%	327

## Indivíduo 11 – N3Z

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	5	100,0%	0	,0%	5
lugar	7	100,0%	0	,0%	7
mulher	5	100,0%	0	,0%	5
qualquer	4	100,0%	0	,0%	4
computador	4	100,0%	0	,0%	4
maior	8	100,0%	0	,0%	8
mar	9	100,0%	0	,0%	9
governador	4	100,0%	0	,0%	4
jogador	5	100,0%	0	,0%	5
dólar	3	75,0%	1	25,0%	4
para	54	100,0%	0	,0%	54
ele	33	100,0%	0	,0%	33
ela	26	100,0%	0	,0%	26
eles	13	100,0%	0	,0%	13
óculos	5	83,3%	1	16,7%	6
era	17	100,0%	0	,0%	17
dinheiro	13	100,0%	0	,0%	13
feira	3	100,0%	0	,0%	3
agora	17	100,0%	0	,0%	17
Lula	5	100,0%	0	,0%	5
problema	9	100,0%	0	,0%	9
outro	8	100,0%	0	,0%	8
quatro	6	100,0%	0	,0%	6
sempre	12	92,3%	1	7,7%	13
primeiro	5	100,0%	0	,0%	5
Brasil	10	100,0%	0	,0%	10
presidente	7	100,0%	0	,0%	7
três	4	100,0%	0	,0%	4
grande	9	100,0%	0	,0%	9
livro	8	100,0%	0	,0%	8
<b>TOTAL</b>	<b>318</b>	<b>99,1%</b>	<b>3</b>	<b>,9%</b>	<b>321</b>

## Indivíduo 12 – N3G

PALAVRAS	Cancelamento				total
	não-cancela		cancela		
	n	%	n	%	
melhor	3	60,0%	2	40,0%	5
lugar	10	76,9%	3	23,1%	13
mulher	7	100,0%	0	,0%	7
qualquer	12	100,0%	0	,0%	12
computador	8	100,0%	0	,0%	8
maior	4	100,0%	0	,0%	4
mar	5	83,3%	1	16,7%	6
governador	4	100,0%	0	,0%	4
jogador	3	75,0%	1	25,0%	4
dólar	4	100,0%	0	,0%	4
para	39	92,9%	3	7,1%	42
ele	59	93,7%	4	6,3%	63
ela	18	94,7%	1	5,3%	19
eles	25	100,0%	0	,0%	25
óculos	5	100,0%	0	,0%	5
era	11	100,0%	0	,0%	11
dinheiro	7	100,0%	0	,0%	7
feira	7	100,0%	0	,0%	7
agora	14	93,3%	1	6,7%	15
Lula	8	100,0%	0	,0%	8
problema	5	100,0%	0	,0%	5
outro	4	33,3%	8	66,7%	12
quatro	5	100,0%	0	,0%	5
sempre	5	71,4%	2	28,6%	7
primeiro	4	100,0%	0	,0%	4
Brasil	10	100,0%	0	,0%	10
presidente	4	100,0%	0	,0%	4
três	7	100,0%	0	,0%	7
grande	6	100,0%	0	,0%	6
livro	5	83,3%	1	16,7%	6
<b>TOTAL</b>	<b>308</b>	<b>91,9%</b>	<b>27</b>	<b>8,1%</b>	<b>335</b>

## APÊNDICE G

### Tabelas – palavras e indivíduos do estudo longitudinal

Abaixo, apresentamos tabelas com os percentuais de cancelamento de cada um dos 2 indivíduos do estudo longitudinal, N3Z e L2R, quanto às 16 palavras avaliadas em tal estudo.

#### Indivíduo – N3Z

Palavras	Cancelamento				Total
	não-cancela		cancela		
	n	%	n	%	
agora	26	100,0%	0	,0%	26
Brasil	6	100,0%	0	,0%	6
ela	106	100,0%	0	,0%	106
ele	175	98,9%	2	1,1%	177
eles	47	97,9%	1	2,1%	48
era	100	100,0%	0	,0%	100
grande	24	100,0%	0	,0%	24
lugar	8	100,0%	0	,0%	8
maior	6	100,0%	0	,0%	6
melhor	6	100,0%	0	,0%	6
mulher	9	100,0%	0	,0%	9
outro	11	100,0%	0	,0%	11
para	171	96,1%	7	3,9%	178
primeiro	5	100,0%	0	,0%	5
sempre	4	44,4%	5	55,6%	9
três	13	100,0%	0	,0%	13
TOTAL	717	98,0%	15	2,0%	732

#### Indivíduo – L2R

Palavras	Cancelamento				Total
	não-cancela		cancela		
	n	%	n	%	
agora	53	71,6%	21	28,4%	74
Brasil	9	100,0%	0	,0%	9
ela	258	75,0%	86	25,0%	344
ele	160	75,5%	52	24,5%	212
eles	25	32,1%	53	67,9%	78
era	47	92,2%	4	7,8%	51
grande	7	100,0%	0	,0%	7
lugar	7	70,0%	3	30,0%	10
maior	6	100,0%	0	,0%	6
melhor	11	73,3%	4	26,7%	15
mulher	8	80,0%	2	20,0%	10
outro	20	57,1%	15	42,9%	35
para	163	58,0%	118	42,0%	281
primeiro	10	90,9%	1	9,1%	11
sempre	5	55,6%	4	44,4%	9
três	21	100,0%	0	,0%	21
TOTAL	810	69,1%	363	30,9%	1173