Ana Paula da Silva Huback

Efeitos de freqüência nas representações mentais

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

Área de concentração: Lingüística
Linha de pesquisa: D (Organização Sonora da Comunicação Humana)
Orientadora: Profª. Drª. Thaïs Cristófaro-Silva

Belo Horizonte
Faculdade de Letras da UFMG
2007

Livros Grátis

http://www.livrosgratis.com.br

Milhares de livros grátis para download.

Universidade Federal de Minas Gerais
Faculdade de Letras
Programa de Pós-Graduação em Letras: Estudos Lingüísticos

Tese intitulada "Efeitos de freqüência nas representações mentais", de autoria da doutoranda Ana Paula da Silva Huback, aprovada pela banca examinadora constituída pelos seguintes professores:

_____

Profª. Drª. Thaïs Cristófaro-Silva – FALE/UFMG – Orientadora

_____

Profª. Drª. Jania Ramos – FALE/UFMG

_____

Prof. Dr. Fabio Alves – FALE/UFMG

_____

Profª. Drª. Eleonora Cavalcante Albano – Unicamp

_____

Profª. Drª. Maria Marta Pereira Scherre – UnB

Belo Horizonte, 30 de março de 2007.

Avenida Antônio Carlos, 6627 – Belo Horizonte – MG – 31270-901 – Brasil
Tel.: (31)3499-5492 – Fax: (31)3499-5113

Agradecimentos

Se eu decidisse mencionar todos os amigos e professores que me auxiliaram na elaboração deste trabalho, certamente o capítulo de agradecimentos seria o maior desta tese... Então, prefiro citar estritamente apenas aquelas pessoas que, nos quatro anos de doutorado, contribuíram para a conclusão desta etapa da minha vida. Devo agradecer, principalmente:

- À minha orientadora, Profª. Drª. Thaïs Cristófaro-Silva, pela assistência acadêmica imprescindível, lendo e comentado os originais da tese, e também pelas discordâncias necessárias, que me ajudaram a amadurecer pontos de vista e contribuíram para que pudéssemos consolidar uma amizade verdadeira, fundamentada no respeito e na admiração;
- À minha orientadora estrangeira, Profª. Emérita Joan Bybee, pelo excelente ano de convivência e pesquisa em New Mexico, EUA, pela disponibilidade em colaborar com minha pesquisa e pelo exemplo de inteligência, profissionalismo e ética;
- Às professoras Eleonora Albano e Jânia Ramos, pelo redirecionamento da tese no Exame de Qualificação;
- À professora Marta Scherre, pela leitura minuciosa da tese e pelas excelentes contribuições para o aperfeiçoamento desta pesquisa;
- Ao CNPq, pela concessão das bolsas de estudos que possibilitaram minha permanência e pesquisa em New Mexico, e a conclusão do doutorado no Brasil;
- À Raquel Márcia, amiga de valor e caráter inestimáveis, pelo exemplo de bondade, solidariedade e sensatez, pela visita nos Estados Unidos, pela "pousada" em tantas idas a BH, pelas conversas intermináveis e risadas inevitáveis, e por toda ajuda "logística" solucionando questões que eu, à distância, não poderia resolver;
- À Daniela Oliveira, pela delicadeza e presteza em ajudar sempre que necessário;
- Aos amigos Fabinho, Gisele e Marlos, pela ajuda na minha "caça aos informantes";
- À minha família, por ter me ensinado o caminho de volta à minha casa;
- À minha irmã Patrícia e meu cunhado Leandro, pela ajuda, companhia e distração nos momentos de maior estresse, e pelo quarto sempre à minha disposição;

- À minha sobrinha Luiza, por me distrair da minha tese e por dizer que o plural "pastel" é "pastels";
- Ao Alan, pelas aulas e assessoria de estatística;
- Ao Alex, Ricky e Valico, pelo companheirismo em New Mexico e pelas gargalhadas que tornaram o deserto menos árido e mais animado;
- À Gisele, pela identidade e cumplicidade que transcendem os limites desta vida;
- Ao Sandrinho, por toda a ajuda nos problemas técnicos da tese, mas, principalmente, pelos 17 anos de confiança mútua;
- À Mag, pelas horas de amizade "virtual" e pelo amor real, numa relação que nem o tempo nem a distância puderam apagar;
- Ao Claudio, pelas caronas e pela companhia nos almoços e lanches nos intervalos das idas e vindas a/de BH;
- Aos amigos que deixei pelo caminho e que não vejo mais com a freqüência que gostaria. Trago todos comigo, naquilo de vocês que faz parte da minha história de vida.

Resumo

Esta tese analisa os grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u* do Português Brasileiro. Nesses grupos de palavras, observa-se, sobretudo em situações informais de comunicação, a utilização de formas de plurais diferentes daquelas previstas pela gramática normativa do Português Brasileiro, como "escrivões" em vez de "escrivães", "sals" no lugar de "sais", ou "degrais" por "degraus". Através de experimentos realizados com falantes nativos brasileiros, argumenta-se que está havendo uma reorganização desses grupos de plurais, devido, sobretudo, a efeitos de freqüência de tipo (paradigma lingüístico) e de ocorrência (repetição dos itens léxicos). Diferentes modelos de léxico mental são discutidos, a fim de verificar qual deles seria capaz de justificar melhor a distribuição dos dados para essas classes de plurais. Como conclusão, defende-se a existência de um léxico mental plástico, remodelado a partir do uso da língua em situações concretas de comunicação. Tal concepção é compatível com o Modelo de Redes (BYBEE, 1995, 2001) e com a Teoria dos Exemplares (PIERREHUMBERT, 2000, 2001a, 2001b).

Abstract

This dissertation analyzes the Brazilian Portuguese plural groups ending in *–ão*, *–l* and *–u* final diphthong. Among these word groups, especially in informal situations, it is observed that some people adopt one plural form that is different from the one suggested by Brazilian Portuguese normative grammars, such as "escrivões" instead of "escrivães" ("notaries public"), "sals" instead of "sais" ("salt" plural), or "degrais" instead of "degraus" ("steps").

Through different experiments carried out with Brazilian native speakers, it is argued that a reorganization of these plural groups is happening, due to type frequency (linguistic paradigm) and token frequency (repetition of lexical items). Moreover, different models of mental lexicon are discussed, in order to verify which ones would best justify the data distribution for these plural classes. As a conclusion, we propose the existence of a plastic mental lexicon, remodeled based on the way language is used in daily communication. This conception is compatible with the Network Model (BYBEE, 1995, 2001) and with the Exemplar Theory (PIERREHUMBERT, 2000, 2001a, 2001b).

SUMÁRIO

CAPÍTULO 1 ..... **19**

INTRODUÇÃO 19

CAPÍTULO 2 ..... **27**

GRUPOS DE PLURAIS 27

2.1 Concordância nominal no português brasileiro 28

2.2 Plural das palavras terminadas em *–ão* 30

2.2.1 A etimologia das palavras terminadas em *–ão* 31

2.2.2 Plural das palavras terminadas em *–ão* nas gramáticas da língua portuguesa. 35

2.2.3 A morfologia do morfema *–ão* 39

2.3 Plural das palavras terminadas em *–l* e ditongo em *–u* 44

2.3.1 A etimologia das palavras terminadas em *–l* e ditongo em *–u* 48

2.3.2 A morfologia das terminações *–l* e ditongo em *–u* 50

2.4 Conclusões 51

CAPÍTULO 3 ..... **53**

REVISÃO DE LITERATURA 53

3.1 Análises anteriores sobre processos de pluralização do português brasileiro 54

3.1.1 A abordagem de Guy (1981) 56

3.1.2 A abordagem de Scherre (1988) 59

3.1.3 Outras abordagens sobre a concordância nominal de número no português brasileiro 67

3.2 Análise crítica da literatura lingüística sobre concordância nominal de número e grupos de plurais do português brasileiro 72

3.3 O papel da freqüência na concordância nominal de número do português brasileiro 75

3.4 Conclusões 83

CAPÍTULO 4 ..... **85**

QUADRO TEÓRICO 85

4.1 Conexionismo 87
4.2 Teoria de Palavras e Regras 102
4.3 Modelo de Redes 114
4.4 Modelo de Exemplares 132
4.5 Conclusões 148

CAPÍTULO 5 149

METODOLOGIA 149
5.1 Metodologia adotada para a coleta de dados 150
5.1.1 Dados de fala espontânea 150
5.1.2 Dados de fala não-espontânea 156
5.2 Corpora do PB consultados 167
5.2.1 O Corpus NILC/São Carlos 168
5.2.2 O Corpus LAEL 170
5.2.3 O Corpus ASPA 171
5.3 A opção pelo Corpus NILC/São Carlos 172
5.4 Variáveis analisadas para cada grupo de plurais 178
5.4.1 Análise das palavras terminadas em *–ão* 178
5.4.2 Análise das palavras terminadas em *–l* e ditongo em *–u* 184
5.5 Fatores sociais 189
5.6 A realização do trabalho de campo 192

CAPÍTULO 6 199

ANÁLISE DE DADOS 199
6.1 Resultados para os plurais terminados em *–ão* 199
6.2 Resultados para os plurais terminados em *–l* 219
6.3 Resultados para os plurais terminados em ditongo em *–u* 234
6.4 Resultados para o "Teste de Reação" 244
6.5 Discussão geral dos resultados 255
6.6 Conclusões 266

CAPÍTULO 7 267

CONCLUSÕES 267

REFERÊNCIAS 273

ANEXOS **291**

ANEXO A 291

Termo de Consentimento Livre e Esclarecido 291

MATERIAL UTILIZADO NOS EXPERIMENTOS 295

1. Conjunto de Figuras 295
2. Leitura de Frases 303
3. Teste de Reação 311

TESTE SOBRE GLIDIFICAÇÃO 313

1. Leitura de Frases 313
2. Plurais 314

ANEXO B – TABELAS REFERENTES AO "TESTE DE REAÇÃO" 315

Índice de Ilustrações

Quadro 1: Formas de singular e plural existentes no latim....31
Quadro 2: Evolução das formas do singular latino até o português *–ão* 33
Quadro 3: Evolução das formas de plural em *–ão* do latim ao português....33
Tabela 1: Processos morfofonológicos de formação de plural e preservação do plural 58
Tabela 2: Processos morfofonológicos de formação do plural de todos os dados dos adultos 61
Tabela 3: Análise final de processos e tonicidade dos substantivos, adjetivos e categorias substantivadas dos adultos por escolarização 64
Tabela 4: Comparação entre os plurais regulares, regulares em *–ão* e plurais em *–ão* com mudança de radical 66
Tabela 5: Freqüência dos tipos terminados em *–ão* (plural) no ASPA....77
Tabela 6: Freqüência dos tipos terminados em *–ão* no Novo Dicionário Aurélio Eletrônico ...78
Tabela 7: Quantidade de itens terminados em *–ão* no corpus do português do século XIV.....79
Tabela 8: Freqüência dos tipos terminados em *–l* (singular e plural) no ASPA 80
Tabela 9: Freqüência dos tipos terminados em ditongo em *–u* (singular e plural) no ASPA 80
Tabela 10: Freqüência dos tipos terminados em *–l* no Novo Dicionário Aurélio Eletrônico ...81
Tabela 11: Freqüência dos tipos terminados em ditongo em *–u* no Novo Dicionário Aurélio Eletrônico 81
Figura 1: A estrutura básica do Modelo Conexionista de Rumelhart e McClelland (1986) 89
Figura 2: Ilustração sobre a Teoria de Palavra e Regras....107
Figura 3: Diagrama sobre a estrutura e o funcionamento da gramática e do léxico mental 109
Esquema 1: Interconexões fonéticas para palavras terminadas em *–ão* no singular e no plural 120
Figura 4: O espaço acústico/articulatório das vogais do inglês 138
Figura 5: Nuvem de exemplares 139
Figura 6: Itens em competição na nuvem de exemplares 140
Tabela 12: Palavras selecionadas para o "Conjunto de Figuras" (freqüência plural) 158
Tabela 13: Palavras selecionadas para a "Leitura de Frases" (freqüência plural)....162
Tabela 14: Palavras selecionadas para o "Teste de Reação" (freqüência plural)....165
Tabela 15: Dados sobre o Corpus NILC/São Carlos 168
Tabela 16: Totais de palavras dos registros e tipos de textos do Corpus LAEL....170

Tabela 17: Comparação entre as cinco palavras mais freqüentes de plural em *–ãos* e *–ães* nos corpora NILC/São Carlos e LAEL (Escrita e Fala) 174
Tabela 18: As palavras mais freqüentes de cada grupo de plural, no LAEL Fala e no LAEL Escrita 177
Tabela 19: Distribuição dos informantes de acordo com fatores sociais 193
Tabela 20: Resultados gerais para os plurais terminados em *–ão* 200
Tabela 21: Efeito da freqüência de ocorrência nos plurais em *–ão* 203
Tabela 22: Palavras adotadas nos experimentos sobre plurais em *–ão* 204
Tabela 23: Efeito do plural etimológico nos plurais em *–ão* 205
Tabela 24: Efeito da freqüência de ocorrência nas variações de plural etimológico *–ões* 206
Tabela 25: Resultados gerais para palavras etimológicas em *–ãos* e *–ães* 208
Tabela 26: Efeito da palavra nos plurais em *–ão* 209
Tabela 27: Reanálise do efeito da palavra nos plurais em *–ão* 210
Tabela 28: Efeito da quantidade de plurais possíveis nos plurais em *–ão* 212
Tabela 29: Correlação entre freqüência de ocorrência e quantidade de plurais possíveis para plurais em *–ão* 213
Tabela 30: Efeito do número de sílabas nos plurais em *–ão* 214
Tabela 31: Efeito da freqüência de ocorrência nos plurais em *–ão* 215
Gráfico 1: Efeito da freqüência de ocorrência nos plurais em *–ão* 215
Tabela 32: Efeito da tonicidade nos plurais em *–ão* 217
Tabela 33: Efeito da faixa etária nos plurais em *–ão* 218
Tabela 34: Resultados gerais para os plurais em *–l* 219
Tabela 35: Efeito da palavra nos plurais em *–l* 221
Tabela 36: Reanálise do efeito da palavra nos plurais em *–l* 222
Tabela 37: Efeito da estrutura morfológica nos plurais em *–l* 223
Tabela 38: Reanálise do efeito da estrutura morfológica nos plurais em *–l* 224
Tabela 39: Efeito da vogal precedente nos plurais em *–l* 225
Tabela 40: Reanálise do efeito da vogal precedente nos plurais em *–l* 226
Tabela 41: Efeito do número de sílabas nos plurais em *–l* 228
Tabela 42: Efeito da freqüência de ocorrência nos plurais em *–l* 231
Tabela 43: Reanálise do efeito da freqüência de ocorrência nos plurais em *–l* sem a palavra "gol" 232
Gráfico 2: Efeito da freqüência de ocorrência nos plurais em *–l* sem a palavra "gol" 232

Tabela 44: Efeito da escolaridade nos plurais em *–l*.....233
Tabela 45: Resultados gerais para os plurais em ditongo em *–u*.....234
Tabela 46: Efeito da palavra nos plurais em ditongo em *–u*.....236
Tabela 47: Reanálise do efeito da palavra nos plurais em ditongo em *–u* 237
Tabela 48: Efeito da vogal precedente nos plurais em ditongo em *–u* 238
Tabela 49: Efeito da estrutura morfológica nos plurais em ditongo em *–u* 239
Tabela 50: Efeito do número de sílabas nos plurais em ditongo em *–u* 239
Tabela 51: Efeito da freqüência de ocorrência nos plurais em ditongo em *–u*.....240
Gráfico 3: Efeito da freqüência de ocorrência nos plurais em ditongo em *–u* 241
Tabela 52: Efeito da faixa etária nos plurais em ditongo em *–u* 242
Tabela 53: Efeito da escolaridade nos plurais em ditongo em *–u*.....243
Tabela 54: Efeito do gênero nos plurais em ditongo em *–u*.....243
Tabela 55: Efeito da palavra no "Teste de Reação" para plurais em *–ão*.....245
Tabela 56: Efeito do tipo de plural no "Teste de Reação" para plurais em *–ão* 246
Tabela 57: Efeito da freqüência de ocorrência no "Teste de Reação" para plurais em *–ão*.....247
Tabela 58: Efeito da palavra no "Teste de Reação" para plurais em *–l*.....248
Tabela 59: Efeito da tonicidade no "Teste de Reação" para plurais em *–l* 249
Tabela 60: Efeito da palavra no "Teste de Reação" para plurais em ditongo em *–u*.....252
Tabela 61: Efeito da freqüência de ocorrência no "Teste de Reação" para plurais em ditongo em *–u*.....254
Tabela 62: Efeito do número de sílabas no "Teste de Reação" para plurais em *–l*.....315
Tabela 63: Efeito da vogal precedente no "Teste de Reação" para plurais em *–l*.....315
Tabela 64: Efeito da estrutura morfológica no "Teste de Reação" para plurais em *–l*.....316
Tabela 65: Efeito da estrutura morfológica no "Teste de Reação" para plurais em ditongo em *–u* 316
Tabela 66: Efeito da vogal precedente no "Teste de Reação" para plurais em ditongo em *–u*317
Tabela 67: Efeito da escolaridade no "Teste de Reação" para plurais em ditongo em *–u*.....317

CAPÍTULO 1

Introdução

A presente tese propõe uma análise sincrônica de três grupos de plurais do Português Brasileiro (doravante PB), a saber:

- Palavras terminadas em *–ão* no singular e que, no plural, são flexionadas através dos morfemas *–ões*, *–ãos* ou *–ães*: "leão"/"leões", "cidadão"/"cidadãos", "alemão"/"alemães";
- Palavras terminadas em *–l* no singular e que, no plural, flexionam-se através da adição do morfema *–is* (com alterações morfofonológicas): "jornal"/"jornais", "anel"/"anéis", "funil"/"funis", "sol"/"sóis", "azul"/"azuis";
- Palavras terminadas em ditongo em *–u*[1] no singular e que, no plural, recebem o morfema *–s*: "degrau"/"degraus", "chapéu"/"chapéus".

Tais grupos de plurais apresentam peculiaridades dignas de nota. Com relação às palavras terminadas no ditongo nasal *–ão*, observamos que a maioria delas é pluralizada com a transformação de *–ão* em *–ões* (Cf. corpora do PB consultados, no Capítulo 3, "Revisão de Literatura"), como em "avião"/"aviões", "leão"/"leões". No entanto, existe um pequeno grupo de palavras dessa classe que se pluraliza com os morfemas *–ãos* ou *–ães*, como "irmão"/"irmãos" e "pão"/"pães". Dentro desse grupo, há itens que, segundo gramáticas do PB[2], apresentam dois ou até três plurais distintos, como "verão" ("verões" e "verãos") e "vulcão" ("vulcões", "vulcãos" e "vulcães"). Nesses casos, há indícios (apresentados posteriormente) de que, em geral, os falantes preferem a forma terminada em

[1] Ao adotarmos a nomenclatura "palavras terminadas em ditongo em *–u*", estamos nos referindo ao grupo de palavras que termina em vogal + semivogal [w], como "degrau", "europeu", "troféu". Não queremos confundir essa classe com a de itens terminados em vogal [u], como "caju" ou "urubu", por isso não adotamos o título "palavras terminadas em *–u*".
[2] Cf. ALMEIDA, 1997; CUNHA; CINTRA, 1985; ROCHA LIMA, 1984.

–ões[3], por isso palavras como "cidadões" (em vez de "cidadãos") e "guardiões" (em vez de "guardiães") são esporadicamente adotadas.

Os dois outros grupos de palavras que serão destacados nesta tese são formados por palavras terminadas em *–l* e *–u* (quando este forma ditongo). Itens como "degrau" e "troféu" deveriam, de acordo com a gramática normativa, ser pluralizados como "degraus" e "troféus", já que a regra prevê que palavras terminadas em vogal formam seu plural através do acréscimo de *–s*; no entanto, algumas vezes as variações "degrais" e "troféis" também são noticiadas. Uma possível explicação para a existência de tais formas pode ser o fato de que, na maioria dos dialetos do PB, houve a perda de distinção fonética entre *–l* e *–u* em fim de sílaba (Cf. CRISTÓFARO-SILVA, 2002, p. 63), tanto que itens como "mal" e "mau" têm a mesma pronúncia, apesar de serem grafados com letras diferentes. Então, por causa disso, é provável que os falantes estejam aplicando a regra de pluralização em *–l* (acréscimo de *–is*) para os itens terminados em ditongo em *–u* (cujo plural seria através do acréscimo de *–s*), como nos casos exemplificados acima.

Esses três grupos de plurais foram escolhidos porque, embora haja regras determinadas de pluralização para cada um deles (Cf. Capítulo 2, "Grupos de Plurais"), observa-se que, na linguagem corrente, casos de aplicação de um plural pelo outro ocorrem ocasionalmente. O Corpus LAEL Fala[4], por exemplo, registra seis ocorrências de "cidadões", duas de "guardiões" e uma de "chapéis"[5]. Tais casos demonstram que, a despeito das regras previstas pela gramática normativa, algumas variações ocorrem nas classes de plurais estudadas nesta tese. Nossa análise, portanto, propõe uma investigação da variabilidade sincrônica desses grupos específicos de plurais do PB.

Na literatura lingüística sobre o PB, o tema concordância nominal de número já foi analisado por alguns autores, principalmente Guy (1981) e Scherre (1988). No entanto, uma característica comum a essas pesquisas é que elas utilizaram como foco de análise a Teoria da Variação e Mudança Lingüística, conforme proposta por Labov (1972). Uma das grandes vantagens dessa abordagem é o fato de ela buscar correlatos entre fenômenos lingüísticos e peculiaridades sociais da comunidade de fala investigada. No entanto, como essa teoria tem seu ponto central no aspecto social das variações e mudanças lingüísticas, acaba por abster-se da análise das operações mentais envolvidas no armazenamento e acesso às palavras no léxico mental do falante. Portanto, trabalhos

---

[3] Cf. CAMARA JÚNIOR, 1972; NUNES, 1945; PARREIRA; PINTO, 1985; SAID ALI, 1964.
[4] Mais detalhes sobre os corpora adotados nesta pesquisa serão mencionados no Capítulo 5, "Metodologia".
[5] Os plurais recomendados pela gramática normativa do PB seriam "cidadãos", "guardiães" e "chapéus".

anteriores sobre concordância nominal no PB, a exemplo de Guy (1981) e Scherre (1988), focalizaram sua análise nas variáveis lingüísticas e extralingüísticas envolvidas nesse fenômeno, excluindo aspectos relacionados às representações mentais de informações lingüísticas no léxico mental do falante.

Outro ponto que deve ser mencionado com relação a análises anteriores sobre concordância nominal no PB é o fato de que o tema principal de investigação foi presença ou ausência das marcas formais de pluralização. Estudavam-se, portanto, os fatores favorecedores e desfavorecedores da marcação de plural. Questões como a variabilidade de plurais em grupos específicos, como *–ão*, *–l* e ditongo em *–u* (conforme mencionado acima) não mereceram tratamento diferenciado nas análises anteriores, porque esse não era o objetivo central de tais pesquisas. Assim, os trabalhos de Guy (1981) e Scherre (1988) examinam todos os processos de pluralização no PB, com a intenção de observar as variáveis lingüísticas e extralingüísticas que determinariam a presença ou a ausência da marcação de plural. No entanto, fenômenos peculiares a classes específicas de plural não foram analisados de forma mais aprofundada.

A partir dessa breve análise crítica da literatura lingüística sobre a concordância nominal no PB[6], percebemos a existência de duas lacunas: i) A falta de uma análise específica sobre fenômenos de pluralização que ocorrem apenas em grupos particulares e ii) A investigação de tais grupos de plurais e seu correlato com o processamento de informações lingüísticas no léxico mental do falante.

Dentro desse contexto, esta tese apresenta uma proposta inovadora na literatura lingüística sobre concordância nominal de número do PB. Essa análise é fundamentada em dois pontos principais: 1) Em primeiro lugar, com relação ao objeto de estudo, centramos nossa investigação em apenas três grupos específicos de plurais, por considerarmos que a pluralização como um todo já dispõe de análises bastante aprofundadas (Cf. GUY, 1981; SCHERRE, 1988) e por percebermos que variações particulares noticiadas apenas nesses grupos de plurais (em *–ão*, *–l* e ditongo em *–u*) não tiveram, ainda, nenhuma análise específica na literatura lingüística; 2) Em segundo lugar, quanto ao foco teórico, diferentemente de análises anteriores, em vez de adotar a Teoria da Variação e Mudança Lingüística, nossa tese propõe uma análise de tais grupos de plurais à luz das teorias de representações mentais, que investigam como o léxico mental dos

[6] No Capítulo 3, "Revisão de Literatura", faremos, de forma detalhada, uma resenha sobre os principais trabalhos que a literatura lingüística apresenta sobre a pluralização no PB.

falantes armazena, categoriza e acessa informações lingüísticas. Nos parágrafos subseqüentes, faremos um breve resumo acerca de tais teorias, procurando interligar o objeto de estudo desta pesquisa com as contribuições teóricas sobre o léxico mental.

Como foco teórico para a pesquisa, adotamos o Modelo de Redes (BYBEE, 1985, 1995, 2001) e a Teoria dos Exemplares (PIERREHUMBERT, 2000, 2001a, 2001b). Tais abordagens postulam que a mente humana categoriza e armazena informações lingüísticas e não-lingüísticas de forma análoga, ou seja, palavras, sons (estruturas lingüísticas em geral) são estocadas da mesma forma como o são cores, animais, plantas e outros objetos do mundo cognoscível. Com base em relações de semelhança e similaridade, o cérebro humano processa, através dos mesmos princípios básicos, as diversas informações advindas do mundo externo. O Modelo de Redes e a Teoria dos Exemplares propõem a existência de um léxico mental organizado a partir do uso concreto das formas lingüísticas no cotidiano do falante. Essas propostas decorrem de modelos análogos com concepção de uso (Cf. LANGACKER, 1987). Segundo tais teorias, o léxico mental apresenta grande capacidade de memória, sendo capaz, portanto, de armazenar detalhamento fonético sobre as palavras, além de informações de ordem social e pragmática acerca do contexto situacional em que determinado item léxico foi utilizado (Cf. JOHNSON, 1997). Além disso, a unidade básica de armazenamento lexical é a palavra inteira, conforme presente na comunicação diária, porque ela é o elemento mínimo a partir do qual se constrói o significado. Nesse tipo de teoria, a freqüência com que certas estruturas lingüísticas ocorrem desempenha papel importante na categorização, armazenamento e emergência de novos padrões dentro da língua. Os itens léxicos são organizados em redes de similaridades semânticas e fonológicas. Informações morfológicas, tais como a existência de afixos, emergem a partir das relações estabelecidas entre os itens pertencentes a essas redes. As representações mentais são reorganizáveis a partir do uso concreto da língua, por isso questões como o uso de uma palavra (freqüência de ocorrência) ou de um padrão lingüístico (freqüência de tipo) desempenham papel crucial na categorização e organização do léxico mental, bem como no acesso aos itens lexicais. Em linhas gerais, pode-se afirmar que palavras que são muito usadas adquirem certa autonomia lexical, por isso podem desligar-se de sua rede e tornarem-se, de certa forma, menos suscetíveis a mudanças; por outro lado, palavras pouco freqüentes são menos disponíveis no léxico mental, por isso algumas vezes acabam sofrendo generalizações em direção a padrões mais freqüentes. Da mesma maneira, quando um determinado paradigma

é bastante recorrente na língua, ele pode se aplicar a novos itens, fazendo com que palavras que originalmente pertenciam a uma determinada classe passem a ser flexionadas a partir de outro paradigma. Dessa forma, o uso que os falantes fazem da linguagem afeta as representações mentais e, conseqüentemente, a estrutura gramatical da língua. Segundo o Modelo de Redes e a Teoria dos Exemplares, a gramática não é uma organização fixa e imutável; ao contrário disso, suas propriedades emergem do uso concreto das estruturas lingüísticas no cotidiano de uma comunidade de fala (Cf. HOPPER, 1998). Portanto, o Modelo de Redes e a Teoria dos Exemplares buscam correlatos entre fenômenos lingüísticos e a forma como o léxico mental categoriza, armazena e acessa informações lingüísticas.

Dentro do escopo teórico e do objeto de estudo delineados acima, podemos afirmar que o objetivo geral desta tese é analisar a interface entre uso lingüístico e representações mentais, verificando como a freqüência das estruturas lingüísticas interfere na organização do léxico mental e nas variações observadas nos grupos de plurais em *–ão*, *–l* e ditongo em *–u* do PB. Com base nos pressupostos teóricos supracitados, argumenta-se que está havendo uma reorganização dessas classes de pluralização, devido, dentre outros fatores, a efeitos de freqüência. Em termos específicos, os objetivos desta tese são os seguintes:

- Comparar diferentes abordagens de léxico mental e verificar como elas poderiam contribuir para a análise dos plurais em *–ão*, *–l* e ditongo em *–u* do PB;
- Observar efeitos de freqüência (de ocorrência ou de tipo[7]) nas variações de plural, verificando se o fato de uma palavra ou padrão lingüístico ser mais ou menos freqüente interfere na sua suscetibilidade a generalizações analógicas;
- Analisar o efeito da freqüência de uso nas representações mentais de informações lingüísticas;
- Identificar em que medida abordagens como o Modelo de Redes e a Teoria dos Exemplares oferecem contribuições para a interpretação e análise das possíveis variações nos grupos de plurais do PB;
- Investigar fatores extralingüísticos, como faixa etária, gênero e escolaridade, que possam interferir nos processos de pluralização do PB.

[7] No Capítulo 4, "Quadro Teórico", conceituamos essas duas análises de freqüência.

Nesta tese, conforme esclarecemos anteriormente, as correntes teóricas adotadas são o Modelo de Redes e a Teoria dos Exemplares; no entanto, para efeito de comparação, outras teorias sobre representações mentais, como o Modelo Conexionista[8] e a Teoria de Palavras e Regras (PINKER, 1991, 1999) serão mencionadas. As justificativas para que tais teorias sejam, também, comentadas nesta tese, são duas: em primeiro lugar, é imprescindível que diferentes concepções sobre léxico mental sejam comparadas a fim de que se compreendam as contribuições de cada uma dessas teorias para a análise dos fenômenos lingüísticos; em segundo lugar, é importante delinear um quadro cronológico que explicite a evolução e as controvérsias entre as teorias sobre léxico mental já propostas na literatura lingüística. Assim, nossa opção teórica para a análise dos dados é pelo Modelo de Redes e pela Teoria dos Exemplares, no entanto, no Capítulo 4, "Quadro Teórico", comentamos o Modelo Conexionista e a Teoria de Palavras e Regras, a fim de propor uma visão mais global sobre as diferentes concepções de léxico mental.

Com o objetivo de analisar os processos morfofonológicos de pluralização do PB (especificamente, nos grupos de plurais em *–ão*, *–l* e ditongo em *–u*) a partir do Modelo de Redes e da Teoria dos Exemplares, elaboramos uma metodologia fundamentada em experimentos. A elicitação de dados de fala espontânea seria desejável, no entanto a baixa ocorrência de palavras desses grupos de plurais (Cf. Capítulos 3, "Revisão de Literatura, e 4, "Metodologia"), inviabilizou essa intenção. Portanto, os dados analisados nesta tese foram coletados a partir de experimentos com o objetivo de estimular os falantes a pronunciarem palavras terminadas em *–ão*, *–l* e ditongo em *–u*, todas em suas formas plurais. A partir de variáveis lingüísticas e extralingüísticas, investigamos quando e por que variações mencionadas anteriormente, como "cidad<u>ões</u>" e "trof<u>éis</u>", ocorreram.

Em termos de organização, esta tese é composta por mais seis capítulos, distribuídos da seguinte maneira:

- No Capítulo 2 ("Grupos de Plurais"), apresentamos informações diacrônicas sobre os percursos evolutivos dos plurais em *–ão*, *–l* e ditongo em *–u* desde o latim até o português moderno;
- No Capítulo 3 ("Revisão de Literatura"), organizamos uma resenha sobre as principais análises que a literatura lingüística já propôs para os plurais do PB e

[8] RUMELHART; McCLELLAND, 1986; DAUGHERTY; SEIDENBERG, 1994; CHRISTIANSEN; CHATER, 2000.

apontamos eventuais lacunas que foram deixadas por esses estudos anteriores. Neste mesmo capítulo, apresentamos uma análise alternativa para os grupos de plurais em *–ão*, *–l* e ditongo em *–u* do PB e mencionamos informações referentes ao papel da freqüência na organização das representações mentais;

- No Capítulo 4 ("Quadro Teórico"), apresentamos um histórico sobre as diferentes concepções de léxico mental já propostas na literatura lingüística e identificamos hipóteses dessas teorias para os grupos de plurais sob análise em nossa tese. Ainda neste capítulo, fazemos opção teórica clara pelo Modelo de Redes e pela Teoria dos Exemplares, justificando tal escolha;
- No Capítulo 5 ("Metodologia"), detalhamos os procedimentos adotados para a coleta de dados, especificando informações sobre os corpora do PB utilizados como consulta, a seleção dos informantes entrevistados e as variáveis lingüísticas e extralingüísticas selecionadas como possivelmente relevantes para o fenômeno;
- No Capítulo 6 ("Análise de Dados"), apresentamos os resultados de nossa pesquisa, procurando explicar a distribuição dos dados através dos pressupostos do Modelo de Redes e da Teoria dos Exemplares, as duas opções teóricas adotadas por esta tese;
- No Capítulo 7 ("Conclusões"), retomamos as principais conclusões alcançadas com esta análise e apontamos possíveis desdobramentos para este trabalho.

Feitas essas considerações sobre os objetivos, foco teórico e organização de nossa tese, passamos ao Capítulo 2, que apresenta as primeiras informações sobre os diferentes grupos de plurais do PB que são objeto de estudo desta pesquisa.

CAPÍTULO 2

Grupos de Plurais

Os Capítulos 2 e 3 desta tese têm como objetivo principal apresentar informações sobre a pluralização dos grupos de palavras terminadas em *–ão*, *–l* e ditongo em *–u* do PB. Basicamente, o presente capítulo expõe informações extraídas de gramáticas e dicionários, enquanto o Capítulo 3 ("Revisão de Literatura") apresenta uma resenha sobre as principais pesquisas lingüísticas já realizadas sobre esses grupos de plurais do PB. Optamos por essa divisão em dois capítulos distintos porque preferimos mencionar, primeiramente, informações de ordem diacrônica e morfológica sobre o fenômeno investigado, para depois contrapor tais elementos com análises lingüísticas já realizadas sobre os plurais investigados neste trabalho.

Nesta tese, analisamos os seguintes grupos de plurais do PB: palavras terminadas em ditongo nasal *–ão* no singular ("avião", "pão"), palavras terminadas em *–l* ("funil", "lençol") e palavras terminadas em ditongo em *–u* ("chapéu", "degrau"). Sendo assim, organizamos o presente capítulo da seguinte forma: a primeira subseção (2.1) apresenta dados gerais sobre a concordância nominal no PB, enquanto as seguintes (2.2 e 2.3) expõem, separadamente, informações sobre os grupos de plurais analisados nesta tese[9]. Ao final deste capítulo, faremos uma breve conclusão (subseção 2.4), retomando os principais pontos sobre os grupos de plurais analisados e conectando tais informações com os tópicos a serem discutidos no capítulo subseqüente.

[9] Comentamos os grupos em *–l* e ditongo em *–u* em uma única subseção, conforme explicado posteriormente.

2.1 Concordância nominal no português brasileiro

No latim, a concordância nominal e verbal obedecia a regras de caso (função sintática na frase), número, gênero e pessoa. Dessas quatro marcas, apenas caso, número e gênero se aplicavam à concordância nominal, visto que a flexão de pessoa era uma marca especificamente relacionada à concordância verbal (Cf. GUY, 1981, p. 89-90). A existência de casos sintáticos permitia que houvesse uma ordem relativamente livre dos termos na oração. Em geral, a única regra observada era com relação ao verbo, que costumava ser colocado no final da frase. Na flexão nominal latina, todos os integrantes de um sintagma seguiam o mesmo padrão de concordância, por isso recebiam marcas de plural para identificar o número, o caso e o gênero. Na evolução para o português moderno, o fenômeno fonológico de perda dos segmentos finais das palavras gerou o apagamento das marcas de caso. Posteriormente, com o desenvolvimento de preposições que substituíam as desinências de caso na identificação sintática das palavras, as declinações foram definitivamente perdidas e a ordem dos vocábulos na oração passou a ser mais fixa. Como resultado dessas transformações, a concordância nominal no PB é baseada em número e gênero. Nesta tese, analisamos a flexão nominal de número em grupos de plurais específicos (terminados em *–ão*, *–l* e ditongo em *–u*).

Com relação à flexão de número dos substantivos, Rocha Lima (1984, p. 73) afirma que há duas possibilidades: o singular, "que se refere a um ente ou um grupo de entes", como "menino", "gato", "rebanho", e o plural, "que denota mais de um ente ou grupo de entes", como "meninos", "gatos", "rebanhos". O mesmo autor (ROCHA LIMA, 1984, p. 73) faz outra observação importante: "Note-se que o singular pode exprimir um ser individual (*o homem que esteve ontem aqui*, *o relógio que comprei*, *um cavalo*), ou uma espécie (*o homem é mortal*, *o relógio serve para marcar as horas*, *o cavalo é ruminante*)."

Nos exemplos citados, um item singular é utilizado para denotar uma classe formada por mais de um elemento ("todos os homens", "todos os relógios", "todos os cavalos"). Essa questão de um item ser usado no singular para indicar genericamente um grupo como um todo será retomada posteriormente, no Capítulo 6, "Análise de Dados".

Rocha Lima (1984, p. 74) também menciona que, em geral, nomes de massa, por não serem contáveis como unidades, conservam-se no singular. É o caso de "água",

"leite", "manteiga", "vinho". Camara Júnior (1991, p. 93) também faz um comentário importante sobre a questão dos nomes de massa:

> A oposição singular : plural permeia todo o conjunto de nomes portugueses. Aplica--se não só ao que em matemática se chama "quantidades descontínuas", vistas como um indivíduo e sua soma, mas também aos "nomes de massa", ou "quantidades contínuas" em que falta a conceituação de indivíduos componentes, como *açúcar*, *farinha*, *ferro*, etc. Neste último caso, a oposição é entre uma única qualidade ou mais de uma qualidade de substância contínua designada, ex.: *açúcares* (vários tipos de açúcar: o refinado, o grosso, o mascavo, etc.).

Kury, Bueno e Oliveira (1977, p. 77) apresentam a regra geral para a concordância no PB: "Os adjetivos, pronomes, artigos e numerais concordam em gênero e número com os substantivos determinados." Vejamos os exemplos abaixo:

Singular: *Meu melhor amigo* viajou ontem.
Plural: *Meus melhores amigos* viajaram ontem.
Singular: *Aquela linda casa mais nova* está à venda.
Plural: *Aquelas lindas casas mais novas* estão à venda.

Com base nos exemplos apresentados, podemos observar o funcionamento da concordância de número prescrita pela gramática normativa do PB. Dentro do sintagma nominal, todos os itens flexionáveis ligados a um determinado núcleo (parte principal do sujeito) concordam em número entre si.

Apesar de a regra básica de concordância prever que todos os integrantes de um mesmo sintagma nominal devem ser formalmente pluralizados (se o núcleo assim o for), existem análises lingüísticas (GUY, 1981; SCHERRE, 1988) que demonstram que, na língua falada, algumas palavras apresentam perda da marca de plural. No português falado, sobretudo em situações espontâneas de comunicação, há uma forte tendência a marcar formalmente o plural somente em itens em posição anterior ao núcleo do sintagma nominal. Itens posteriores ao núcleo são mais propensos ao apagamento das marcas de pluralização e itens anteriores ao núcleo favorecem a presença da marcação de plural (Cf. SCHERRE, 2001a, p. 3). No Capítulo 3, esse assunto sobre a perda da marca de plural será explorado mais profundamente.

As subseções seguintes apresentam informações diacrônicas sobre os grupos de plurais individualmente.

2.2 Plural das palavras terminadas em *–ão*

O PB apresenta algumas palavras que, no singular, terminam em *–ão*, um ditongo nasal. Quando pluralizados, esses itens léxicos dividem-se em dois grupos:

(1) Palavras que se pluralizam através de *–ões*, *–ãos* ou *–ães* (com apenas uma forma de plural para cada item singular):

leão	→	leões	[le'õys]
mão	→	mãos	['mãws]
pão	→	pães	['pãys]

(2) Palavras em que, para um único singular, há duas ou três formas de plural[10]:

guardião	→	guardiães	~	guardiões		
sacristão	→	sacristães	~	sacristãos		
ancião	→	anciãos	~	anciães	~	anciões
vulcão	→	vulcãos	~	vulcães	~	vulcões

O fato de haver palavras com uma, duas ou até três formas possíveis de pluralização gera determinadas controvérsias. Algumas vezes ocorrem casos em que uma palavra que pertence ao grupo 1 acima (com apenas uma forma de plural) adota um plural diferente daquele que seria etimologicamente previsto. É o caso da palavra "cidadão": seu plural seria "cidadãos" (Cf. gramáticas normativas), mas a forma "cidadões" já é registrada

[10] Os exemplos acima foram extraídos do Novo Dicionário Aurélio Eletrônico. Conforme veremos na subseção 2.2.2, as gramáticas normativas do português diferem quanto às formas de plural consideradas corretas para as palavras terminadas em *–ão*. Esse ponto será discutido oportunamente, na referida subseção. Neste momento, queremos apenas salientar o fato de que, no grupo de palavras em *–ão*, existe oposição entre itens que apresentam apenas uma forma de plural (grupo 1 acima) e itens que apresentam dois ou três plurais (grupo 2 acima).

no Corpus LAEL Fala. Esse tipo de fenômeno será investigado mais detalhadamente nesta tese.

Como a quantidade de informações sobre o plural das palavras terminadas em *–ão* é extensa, vamos dividir esta seção em algumas subseções menores, tratando, respectivamente, da etimologia da terminação *–ão* (2.2.1), informações sobre esses plurais em gramáticas do PB (2.2.2) e morfologia da terminação *–ão* (2.2.3).

2.2.1 A etimologia das palavras terminadas em *–ão*

No latim, encontramos a etimologia das três diferentes formas de plural utilizadas para o grupo terminado em *–ão* no PB. Segundo Sequeira (1943), no latim havia, originalmente, as terminações *–anem*, *–anum*, *–onem –onum* e *–udinem* para formas no singular. Cada uma dessas formas apresentava um plural diferente, conforme podemos observar no quadro abaixo:

Quadro 1: Formas de singular e plural existentes no latim

Singular	Plural	Exemplos	
–anem	–anes	panem	panes
–anum[11]	–anos	germanum	germanos
–anum	–anus	manum	manus
–onem	–ones	leonem	leones
–onum	–onos	patronum	patronos
–udinem	–udines	certitudinem	certitudines

Até o século XIV, a situação das formas de singular e plural continuava a mesma; no entanto, a partir desse século, essas terminações sofreram uma série de modificações que resultaram na convergência para uma única forma singular, o *–ão* do PB. Explicando brevemente essas mudanças, primeiramente, a desinência *–em*, relativa ao acusativo singular latino, desapareceu, por causa do enfraquecimento das terminações

[11] A terminação *–anum* apresentava duas formas diferentes de plural: *–anos* e *–anus*, por isso foi registrada duas vezes no Quadro 1.

referentes aos casos latinos e do desenvolvimento gradual das preposições. Em um segundo momento, a nasalidade do segmento /n/ intervocálico foi assimilada pela vogal precedente, fazendo com que o /n/ sofresse síncope. Posteriormente, houve fusão das duas vogais contíguas e, bem mais tarde, a partir de transformações diferentes para cada uma das terminações, as formas singulares se fundiram em *–ão*. Tais transformações podem ser visualizadas através dos exemplos abaixo:

> *pane* > **pãe* > *pã* > *pão*
> *multitudine* > *multidõe* > *multidõ* > **multidõo* > *multidão*
> **coratione* > **coraçõe* > *coraçõ* > *coraçõo* > *coração*
> *manu* > *mão*
> (GUERIOS, 1942, p. 118)

Nunes (1945, p. 226-227) apresenta sua análise para a evolução das formas em *–ão* do português moderno:

> [...] até o século XIV manteve-se entre as terminações *–ao* (*sic*), *–ã*, *–õ* e *–õe* a diferença resultante das suas distintas proveniências, que foram respectivamente *–anu–*, *–ana–*, *–one–* e *–ŭdine–*, depois, *–õe* perdeu o *–e*, como já em tempos pré--literários tinha sucedido a *–ã*, representante de *–ane–*, e a *–õ*, sem dúvida porque o seu plural era igual aos dos nomes da última terminação. Em seguida estas duas vogais nasais, a primeira talvez sob influência de *–ão*, tomaram um *–o* de encosto, resultando de aí os ditongos nasais *–ão* e *–õo* (na pronúncia *ãu* e *õu*). Este último, que ainda persiste nalgumas falas dialectais, ou por dissimilação ou por outra causa que se ignora, tornou-se depois em *–ão*, chegando assim a reduzir-se a uma única forma, *–ão* ou *–am* (como também se escrevia) as quatro antigas, mas continuando a observar-se no plural a diferença que existira no singular.

Nunes (1945) enfatiza que a fusão das formas singulares foi gradual, ocorrendo primeiramente na língua falada, para depois consolidar-se na escrita. Outro ponto importante comentado pelo autor é o fato de que, em textos do século XV, é comum encontrarmos a mesma palavra (do grupo em *–ão*) grafada de duas formas diferentes.

Mattos e Silva (19--, p. 137) adverte o seguinte:

> Nos documentos medievais esse processo de convergência [das formas em *–ão* no singular] aparece reflectido na grafia de textos sobretudo posteriores ao século XIV; no entanto, em documentos literários e não literários do século XIII já se pode (*sic*) rastrear os inícios desse processo.

Portanto, a convergência das formas em *–ão* no singular alcançou seu auge no século XIV, mas, no século XIII, casos isolados dessa fusão já eram noticiados. O quadro abaixo, extraído de Sequeira (1943, p. 48), sumariza a evolução das formas do singular:

Quadro 2: Evolução das formas do singular latino até o português *–ão*

Latim	**Português preliterário (*sic*)**	**Até ao séc. XIV**	**Séc. XIV-XV**	**Desde o séc. XV-XVI**
–one	*–õe	**–õ**		
–udine	–õe	**–õe**	–õo	
–onu	–õo	**–õo**		–ão
–ane	*–ãe	–ã		
–anu[12]	–ão	–ão	–ão	

No quadro acima, percebemos que as terminações que possuíam diferentes formas no singular latino fundiram-se todas no morfema *–ão*, que prevalece até hoje no PB. No entanto, no plural, as transformações sofridas por essas desinências não foram as mesmas, por isso a pluralização até hoje ainda conserva terminações distintas, apesar de haver casos abundantes (formas singulares que admitem mais de um plural) registrados em gramáticas do PB. Segundo Sequeira (1943, p. 48), a evolução das formas de plural foi a seguinte:

Quadro 3: Evolução das formas de plural em *–ão* do latim ao português

Vocábulos latinos	**Plurais latinos**	**Plurais arcaicos**	**Plurais modernos**	**Vocábulos portugueses**
Leones	–ones			*Leões*
Certitudines	–udines	–ões	–ões	*Certidões*
Patronos	–onos			*Padrões*
Panes	–anes	–ães	–ãis	*Pãis* (*sic*)
Germanos	–anos	–ãos	–ãos	*Irmãos*
Manus	–anus			*Mãos*

[12] O referido livro não mencionava a forma *–ano* neste quadro, provavelmente porque o autor já estava considerando a fusão das terminações *–anu* e *–ano*.

[13] A utilização do negrito no quadro acima tem como objetivo ressaltar que os três plurais (*–õ*, *–õe* e *–õo*) fundiram--se em *–õ*.

Parreira e Pinto (1985, p. 161) apresentam um bom sumário sobre a questão da evolução das formas de plural e singular:

> Os substantivos terminados no singular em *–ão* formam o plural de três modos diferentes: *–ãos*, *–ões* e *–ães*. Na passagem do latim para o português verifica-se a queda do *n* intervocálico e a nasalação da vogal tónica e posteriormente a contracção das duas vogais. Enquanto no singular se dá a convergência das três formas latinas diferentes:
>
> *–ão* < *–anum*
> < *–onem*
> < *–anem*
>
> no plural essa convergência não se deu:
>
> *–ãos* < *–anos*
> *–ões* < *–ones*
> *–ães* < *–anes*

Depois dessas transformações fonéticas ocorridas na evolução das formas em *–ão* do português moderno, não existe mais uma utilização consistente do plural etimológico[14] dos itens em *–ão*, assim como não existe, também, uma regra geral que possa ser aplicada a esse grupo. Nunes (1928, p. 164) faz um comentário importante a esse respeito: "Quanto aos [itens] de temas vocálicos, êsses continuaram a formar o respectivo plural consoante a regra própria de tais nomes, isto é, com a adjunção apenas do morfema *–s* ao singular, assim: *cristãos*, *grãos*, *mãos*, *sãos*, *vãos*, etc."

Portanto, com relação às palavras terminadas em *–ão*, apenas o grupo de etimologia em *–ano* conserva a regra geral de pluralização das palavras terminadas em vogal no PB, ou seja, o acréscimo do morfema *–s*. Nos demais grupos (de etimologia em *–ane* ou *–one*), por causa das transformações fonéticas que resultaram na convergência para uma única forma no singular, a regularidade de pluralização a partir do acréscimo do morfema *–s* já não se aplica. Conforme veremos no capítulo "Análise de Dados", o morfema *–ãos*, a despeito de sua pluralização regular, também sofre variação.

Ainda com relação à preservação da etimologia dos itens terminados em *–ão* no singular, uma possível fonte para a recuperação do plural etimológico poderia ser a derivação. Pares de palavras como "capitão"/"capitANIa", "pão"/"pANIficação"; "grão"/"grANUlado" e "mão"/"mANUfaturar" podem fornecer indícios para a manutenção da etimologia. Nos casos de "capitão" e "pão", a derivação nos fornece

[14] Nesta tese, vamos utilizar os termos "plural etimológico" para as formas de plural que eram utilizadas no latim e "plural analógico" para outras formas de plural que se desenvolveram a partir da analogia.

–ANI–, portanto o plural etimológico seria *–anes* (*–ães* no PB); em "grão" e "mão", a derivação nos fornece –ANU–, então o plural etimológico seria *–anus* (*–ãos* no PB). Casos como "leão"/"leONInos" e "abolição"/"aboliciONIsta" recuperam –ONE– no radical, portanto o plural é em *–ões*. A derivação funcionaria como uma boa fonte para a manutenção do plural etimológico, no entanto dois problemas ocorrem: i) Essa derivação não é tão facilmente recuperável, tanto que variações na utilização das diferentes terminações de plural em *–ão* ocorrem. "Corrimão", por exemplo, por ser derivado de "mão", tem plural etimológico em "corrimãos", no entanto a forma "corrimões" já é aceita por gramáticas e dicionários do PB (Cf. ROCHA LIMA, 1984, p. 76); ii) Há casos cujo plural etimológico não pode ser recuperado por derivação, como "verão" – "verANIsta" (a derivação gera –ANI– (*–ães* no PB), mas o plural etimológico é "verãos" e não "*verães"), "órgão" – "orgANIzar" (a derivação também gera –ANI– (*–ães* no PB), mas o plural etimológico é "órgãos", e não "*órgães").

A partir do que foi exposto acima, podemos compreender melhor por que existe controvérsia na pluralização das formas em *–ão* do PB. Em termos de história da língua, as causas básicas desse fato são: 1) As formas de singular em *–ão* fundiram-se em uma só, mas as formas de plural permaneceram diferentes. Por conta disso, o correlato etimológico se perdeu, o que gerou divergências quanto à forma a ser adotada como plural; 2) Por conta da analogia, algumas palavras passaram a ter plural duplo ou tríplice (Cf. FERNANDES, 1947, p. 13-14).

A próxima subseção apresenta os comentários e ponderações da gramática normativa acerca da pluralização de palavras terminadas em *–ão* no singular.

2.2.2 Plural das palavras terminadas em *–ão* nas gramáticas da língua portuguesa

Os gramáticos prescritivos do PB são bastante contundentes na defesa dos plurais adequados a serem usados para a forma *–ão* no singular. Fernandes (1947, p. 15) sugere que, no caso de dúvida quanto ao plural "correto"[15], se recorra ao espanhol, porque,

[15] Nesta tese, o adjetivo "correto", quando referir-se a uma forma lingüística, deve ser entendido como a forma indicada como a mais adequada pela gramática normativa. Não existe, de nossa parte, nenhum julgamento de valor quanto às diferenças em função da variação dialetal.

nessa língua, o /n/ intervocálico não sofreu síncope, então todo o processo de unificação das formas singulares não ocorreu. Por causa disso, temos os correlatos espanhóis e portugueses: "catalanes"/"catalães", "corazones"/"corações", "paganos"/"pagãos". Após fornecer essa instrução para a utilização do plural adequado, o autor prossegue:

> Não havendo possibilidade de tal recurso e desconhecendo-se a origem do vocábulo, então, na falta de pessoa competente que saiba esclarecer o caso, só a leitura constante dos bons escritores poderá decidir com segurança quais as formas de plural que devem ser empregadas. (FERNANDES, 1947, p. 15)

Na opinião de alguns gramáticos, as generalizações de plural são fruto da "confusão que se tem atribuído à boca das massas populares incultas e também ao pouco escrúpulo dos antigos copistas, que se deixavam influenciar pela linguagem do seu meio." (FERNANDES, 1947, p. 13). Portanto, as gramáticas, em geral, condenam as formas de plural analógicas, surgidas, principalmente, a partir da prevalência da forma *–ões*, como veremos a seguir.

Conforme foi afirmado anteriormente, nem mesmo gramáticas conceituadas da língua portuguesa apresentam consenso quanto às palavras que possuem plural abundante (com mais de uma possibilidade de plural para um mesmo item singular) e quanto às formas aceitas como pluralização para cada uma dessas palavras. A fim de comprovar essa afirmação, vamos proceder a uma comparação sobre esse assunto em três gramáticas diferentes: Rocha Lima (1984), Cunha e Cintra (1985) e Almeida (1997).

Rocha Lima (1984, p. 76) cita 13 palavras que apresentam mais de uma forma de plural:

Alão – *alãos*, *alães* e *alões*
Aldeão – *aldeãos* e *aldeões*
Anão – *anãos* e *anões*
Ancião – *anciãos*, *anciães* e *anciões*
Castelão – *castelãos* e *castelões*
Charlatão – *charlatães* e *charlatões*
Corrimão – *corrimãos* e *corrimões*
Deão – *deães*, *deãos* e *deões*
Ermitão – *ermitãos*, *ermitães* e *ermitões*
Guardião – *guardiães* e *guardiões*
Sultão – *sultães* e *sultões*
Verão – *verãos* e *verões*
Vilão – *vilãos* e *vilões*

Ao findar essa lista, Rocha Lima (1984) adverte que as formas "preferíveis" de plural vêm em primeiro lugar. Observe-se que essas flexões correspondem às grafias

etimológicas, tal como eram no latim. Os demais plurais abundantes formaram-se através da analogia popular.

Em Cunha e Cintra (1985, p. 176-177), encontramos outra lista com 16 formas que apresentam plurais duplos ou tríplices:

Alão – *alãos*, *alões* e *alães*
Alazão – *alazães* e *alazões*
Aldeão – *aldeãos*, *aldeões* e *aldeães*
Anão – *anãos* e *anões*
Ancião – *anciãos*, *anciões* e *anciães*
Castelão – *castelãos* e *castelões*
Corrimão – *corrimãos* e *corrimões*
Deão – *deães* e *deões*
Ermitão – *ermitãos*, *ermitães* e *ermitões*
Hortelão – *hortelãos* e *hortelões*
Refrão – *refrães* e *refrãos*
Rufião – *rufiães* e *rufiões*
Sultão – *sultões, sultãos* e *sultães*
Truão – *truães* e *truões*
Verão – *verões* e *verãos*
Vilão – *vilãos* e *vilões*

Cunha e Cintra (1985, p. 177) fazem um comentário importante sobre a forma "corrimão": "*Corrimão*, como composto de *mão*, devia apresentar apenas o plural *corrimãos*; a par desse, existe também *corrimões*, por esquecimento da formação original da palavra." Esse comentário parece indicar que, em uma palavra composta, o falante perde noção das partes originais e, conseqüentemente, pode aplicar a pluralização mais disponível em seu léxico mental, no caso a terminação *–ões*.

Por fim, Almeida (1997) também cita uma lista de 14 palavras que apresentam plural abundante:

Alão – *alãos*, *alães* e *alões*
Alazão – *alazães e alazões*
Alcorão – *Alcorães* e *Alcorões*
Aldeão – *aldeãos*, *aldeães* e *aldeões*
Anão – *anãos*, *anães* e *anões*
Ancião – *anciãos*, *anciães* e *anciões*
Bastião – *bastiães* e *bastiões*
Fuão – *fuãos* e *fuões*
Guardião – *guardiães* e *guardiões*
Peão – *peãos*, *peães* e *peões*
Pião – *piãos, piães* e *piões*
Sacristão – *sacristãos* e *sacristães*
Vilão – *vilãos* e *vilões*
Vulcão – *vulcãos*, *vulcães* e *vulcões*

Se procedermos a uma comparação mais direta das três listas supracitadas, veremos que não existe acordo quanto às palavras apresentadas como suscetíveis a plural

abundante nem quanto às possibilidades de pluralização para cada uma das palavras. Citando apenas um exemplo, Rocha Lima (1984) apresenta a palavra "sultão" com plural duplo ("sultães" e "sultões"). A mesma palavra, em Cunha e Cintra (1985), é registrada com plural tríplice. Em Almeida (1997), esse item não consta na lista dos plurais abundantes. Os exemplos são vários e não nos cabe, aqui, esgotá-los. Com essa comparação, nosso objetivo é demonstrar que, mesmo entre gramáticas normativas do PB, não existe consonância quando o tema é plural das palavras terminadas em *–ão*.

Camara Júnior (1972, p. 68) já noticiava tal discordância:

> From a synchronic viewpoint, the distribution of the three plural types is arbitrary and in fact they are often confused, even to the point of free variation, whether it be accepted by the grammarians (ex: *aldeãos* : *aldeões* : *aldeães* 'peasants') or not (*cidadãos* is accepted while *cidadões* is rejected as the plural of *cidadão* 'citizen').

Outro ponto importante a ser destacado quanto à análise das gramáticas acima é o fato de que, em geral, foram as palavras com plural etimológico *–ãos* ou *–ães* que desenvolveram plural abundante. Exemplos disso são os itens "guardião" e "vulcão", cujas etimologias são em *–ane* e *–ano*, respectivamente. Os dois adotaram plurais em *–ões* ("guardiões" e "vulcões"). O contrário (palavras com plural em *–ões* desenvolverem formas em *–ãos* ou *–ães*) só ocorreu em dois casos nas gramáticas comparadas: os itens "peão" e "pião", que têm sua etimologia em *pedo*, *onis* (plural em *–ões*) e, atualmente, admitem plural tríplice. "Peão", no século XV, era grafado *pyam* (plural em *–ães* ou *–ãos*)[16], e "pião" (o brinquedo) entra na língua portuguesa em 1720. Por causa dessa divergência da grafia do século XV em relação à etimologia *–ões*, esses são exemplos raros de palavras etimológicas em *–ões* que adotaram plural em *–ães* ou *–ãos*. O mais comum é que itens terminados em *–ãos* e *–ães* etimológicos adotem plural em *–ões*, como vimos na lista acima. Por outro lado, é interessante notar, também, que todas as palavras mencionadas, com exceção de "refrão" e "sacristão" (pluralizadas em "refrães"/"refrãos", "sacristães"/"sacristãos"), adotaram o plural em *–ões*, o que parece indicar que essa é a forma preferível na formação de analogias.

[16] O Dicionário Houaiss Eletrônico afirma que a origem de "pião" (o brinquedo) advém de "peão" (o que anda a pé), por comparação com o movimento de um soldado de infantaria. A origem dos dois substantivos é, portanto, a mesma (*pedo*, *onis*), com etimologia *–ões*. A datação do primeiro registro de "pião" na língua portuguesa é 1720.

Ainda com relação à questão da pluralização nas gramáticas normativas, observa-se que essa discordância entre os plurais das palavras terminadas em *–ão* não é um fenômeno recente, pois há indícios de que isso já ocorria desde o português arcaico:

> Quanto àqueles vocábulos, de cada um dos quais se conhecem duas e até três formas de plural, convirá saber que apenas uma é legítima e – o que é pior – nem sempre é esta a mais usada. Esses plurais duplos e tríplices datam do período arcaico da língua e resultaram do pouco cuidado de antigos copistas ou da confusão das massas populares ou ainda de mal fundadas analogias. (FERNANDES, 1947, p. 16)

Nascentes (1953, p. 82), ao analisar variações fonológicas e morfológicas no linguajar carioca, comenta a preferência popular pela forma *–ões* para o plural, em palavras como "capitões" por "capitães" ou "cidadões" por "cidadãos", mas salienta, também, que as formas "capitões" e "cidadões" já eram encontradas no português de Portugal e na obra de Frei Luís Sousa, que data de 1784. Said Ali (1964, p. 60) registra exemplos de plurais não-etimológicos das palavras em *–ão* entre os autores quinhentistas e seiscentistas: "cidadões", "ermitãos" (em vez de "cidadãos" e "ermitães", formas de etimologia latina). Tais informações comprovam, mais uma vez, que as generalizações de plural não são fenômeno recente no PB.

A próxima subseção analisa a morfologia do morfema *–ão*.

2.2.3 A morfologia do morfema *–ão*

Conforme foi mencionado na subseção 2.2.1, o morfema *–ão* do PB tem sua origem nas terminações *–anem*, *–anum*, *–onem*, *–onum* e *–udinem* do latim. Segundo sua origem e evolução, cada uma dessas desinências pode representar um sufixo com diferentes significados no PB. Nesta subseção, analisaremos os sufixos com a terminação *–ão* e as diferentes formas de plurais que esses podem carregar.

Primeiramente, devemos esclarecer que existem palavras, tanto oriundas do latim quanto advindas de outras línguas, que possuem a terminação *–ão*, mas essa é parte do radical do item léxico e, portanto, não representa nenhum sufixo. É o caso, por exemplo, das palavras "agrião", "anão", "avião", "sótão" e "vulcão".

Por sua vez, ao lado dessas palavras cuja terminação *–ão* não representa sufixo, podemos registrar, no PB, alguns significados diferentes para o sufixo *–ão*. Em Coutinho (1981) e Cunha e Cintra (1985), encontramos um bom inventário sobre tais sufixos. Os autores, considerados em conjunto, listam quatro possibilidades diferentes para sufixos em *–ão*. Em geral, tais sufixos são pluralizados através do morfema *–ões*, como nos seguintes casos: 1) "agressão" e "oração" (sufixos *–são* e *–ção*, que denotam ação ou resultado dela), cujos plurais são "agressões" e "orações"; 2) "casacão" e "facão" (sufixo *–ão* aumentativo), cujos plurais são, também, em *–ões*: "casacões" e "facões". Por outro lado, existem, também, outros nomes com sufixo *–ão*, mas que apresentam o plural em *–ãos*. É o caso de "pagão" e "vilão", em que o sufixo denota cargo, origem ou naturalidade e tem sua fonte etimológica em *–anu*, do latim. Essas palavras, portanto, são pluralizadas através do acréscimo do morfema *–ãos*: "pagãos" e "vilãos".

Com relação à freqüência desses sufixos, Campos (1981), analisando corpora do PB e do português de Portugal, afirma que o sufixo *–ção* é o mais freqüente de todos, sendo responsável por 59,43% das palavras sob análise. Segundo a autora, por causa dessa alta freqüência, tal sufixo é bastante produtivo e utilizado sempre que novos substantivos são formados através de verbos ("localizar"/"localização", "movimentar"/"movimentação"). Percebemos, portanto, que uma importante característica a favor da manutenção do plural em *–ões* é o fato de ele possuir status morfológico de sufixo em mais casos que *–ãos* e *–ães*.

A prevalência do plural *–ões* para grande parte das palavras terminadas em *–ão* no singular é comentada e atestada por vários gramáticos. Camara Júnior (1972, p. 68) enfatiza a divergência entre a escolha dos plurais, a predominância da desinência *–ões* e o esforço escolar para fixar a forma etimológica das palavras:

> Although there is significant disagreement on this point among grammarians (as well as in literary usage), a tendency toward fixation of the morphophonemic pattern *–ão* : *–ões* is apparent. This is a consequence of the circumstance that words that formerly had *–om* are much more frequent in the modern language than those that had original *–ão* or *–am*. The *–om* type includes, for example, all action nouns related to a verbal radical (exs: *consolação* 'consolation'; *persuasão* 'persuasion'; *omissão* 'omission'). The leveling process, however, is a very slow one, and common words that derive from an original *–ão* (exs: *irmão*, *mão*) or *–am* (ex: *pão*) remain totally exempt from it. In the schools a valiant attempt to save the three types of plural is under way. Such efforts are based either on the Latin origin of the words in question, their relation to the corresponding Spanish words, which really do have distinct forms (exs: *razón* : *razones*, *hermano* : *hermanos*, *pan* : *panes*), or the preferred usage of certain authors considered 'classic' in the literary language.

Conforme Camara Júnior (1972) afirma, existe esforço das escolas no sentido de ensinar aos alunos as formas etimológicas que devem ser adotadas, no entanto a discordância sobre os plurais adequados para algumas palavras terminadas em *–ão* (no singular) prossegue. Conforme vimos anteriormente, as gramáticas normativas do PB apresentam diferentes palavras com formas variáveis de plurais abundantes, por isso não existe consenso nem mesmo entre os materiais didáticos adotados nas escolas.

Cunha e Cintra (1985, p. 176) também comentam a prevalência do morfema *–ões* sobre *–ãos* e *–ães* na língua falada: "Para alguns substantivos finalizados em *–ão*, não há ainda uma forma de plural definitivamente fixada, notando-se, porém, na linguagem corrente, uma preferência sensível pela formação mais comum, em *–ões*."

Nunes (1945, p. 236-7) salienta a preferência popular pela forma *–ões*:

> Parece até que no povo encontrou mais simpatia a terminação –ÕES, pois para ele os nomes em –ÃO assim fazem geralmente o plural, e não só o povo, a língua culta também, quando quer formar derivados de nomes em –ÃO, procede como se todos êstes terminassem realmente em –OM (...); até a palavra VERÃO, cujo plural devia ser VERÃOS, encosta-se aos mesmos nomes em –OM.

Parreira e Pinto (1985, p. 160-161), após apresentarem uma lista de plurais abundantes, também ressaltam a predominância da terminação *–ões* no PB:

> Preferimos, no entanto, as formas *aldeões*, *corrimões*, *ermitões*, *guardiões*, *verões* e *vilões*, por tenderem a ser as mais usadas. Repare-se que a maioria dos nomes terminados em *–ão* que vieram do latim para o português formam o plural em *–ões*, daí que as palavras novas que entram na língua terminadas em *–ão* tendam a seguir a maioria, formando o plural em *–ões*.

Said Ali (1964, p. 59-60) também menciona a adoção da forma *–ões* para pluralizar quaisquer palavras novas que surjam na língua:

> A regra geral do plural em *–ões* vigora para a linguagem moderna, sendo aplicada naturalmente a quaisquer têrmos novos: *civilizações*, *vagões*, *salões*, *montões*, *felicitações*, *estremeções*, etc. Dêste oceano de substantivos em *–ão*, mal tiramos umas duas dúzias de vocábulos com plural diferente...

A partir de tudo o que foi exposto acima sobre a prevalência da forma *–ões* para pluralizar as palavras em *–ão*, cabe a pergunta: por que essa forma é a mais freqüente das três? Por que justamente ela é adotada como plural analógico? Said Ali (1964, p. 60) apresenta uma possível explicação para o fato:

> Os têrmos em *–ane* e *–anu*, donde se originaram os plurais em *–ães* (português antigo *–ãaes*) e *–ãos* (português antigo *–ãaos*), recebidos do latim, foram mui poucos em comparação da onda de nomes em *–one* com que se enriqueceu o idioma português; e teria havido menos dificuldade em formar o plural dêsses diversos nomes se no singular as terminações *–om*, *–am* e *–ão* houvessem permanecido sempre distintas entre si. Ao contrário disso, principiaram elas cedo a confundir-se na pronúncia, e daí o embaraço não sòmente para o plural de vocábulos de filiação latina, cuja etimologia era obscura ou esquecida, mas ainda para os têrmos que novamente se cunharam ou importaram do estrangeiro.

Dessa forma, parece que o português recebeu, do latim, mais palavras pluralizadas em *–ões* que em *–ãos* ou *–ães*. Fernandes (1947, p. 15-16) também tenta justificar a tendência de adoção do plural em *–ões* no PB:

> Os vocábulos formados dentro da nossa Língua, como *folião*, de *folia* mais *–ão*, *guião*, de *guia* mais *–ão*, etc., com inclusão de todos os que exprimem idéia de aumento (*caixão*, *carão*, *casão*, *espadão*, *facão*, *fardão*, *garrafão*, *janelão*, *mulherão*, *palavrão*, *paredão*, *portão*, *tabuão*, *vagalhão*, etc.), fazem os seus plurais regulares sempre terminados em *–ões*, considerando-se ilegítimas ou resultantes de falsas analogias certas formas, como *foliães*, *guiães* e alguma outra, porventura usada em tal ou tal região.
> A propósito da pluralização em *–ões*, devemos acentuar mais uma vez que é esta a predominante, considerada até como regra geral para os nomes em *–ão*, segundo vários autores. Mais eufónica, tornou-se mais popular e assim vigora na linguagem moderna, aplicando-se de preferência a termos relativamente novos (*civilizações*, *felicitações*, *vagões*, etc.), o que tudo lhe dá particular importância, até pela enormíssima quantidade, quase incontável, desta forma de plurais, ao passo que a custo se reune (*sic*) um escasso meio cento de plurais em *–ãos* e em *–ães* juntos.

Assim, os gramáticos cujos livros foram consultados são unânimes na opinião de que o plural em *–ões* é, realmente, o preferido pelos falantes na flexão de novas palavras e na produção de formas analógicas. Quanto aos plurais em *–ães* e *–ãos*, convém salientar que *–ães*, conforme veremos no capítulo seguinte, é o mais raro, e *–ãos* é um pouco mais freqüente[17]. Quanto a esse último grupo (plurais em *–ãos*), é necessário fornecer mais uma informação:

> Os vocábulos graves ou paroxítonos terminados em *–ão* têm todos o plural ùnicamente em *–ãos*, sejam quais forem as suas origens – regra esta que não tem excepção. Exemplos: *acórdão*, *bênção*, *frângão*, *gólfão*, *lódão*, *médão*, *morángão*, *orégão*, *órfão*, *panórgão*, *rábão*, *sótão*, *támpão*, e *zángão* correspondem, portanto, a plurais terminados em *–ãos*." (FERNANDES 1947, p. 15)

[17] Estatísticas mais precisas sobre esses morfemas serão apresentadas no Capítulo 3, "Revisão de Literatura".

O fato de todos os itens paroxítonos terminados em *–ão* formarem o plural sempre em *–ãos* pode ser uma referência para que não haja variações na pluralização de tais palavras. Como não existem formas paroxítonas pluralizadas em *–ões* ou *–ães*[18], é de se esperar que, nesses itens, o plural em *–ãos* se conserve. Essa tendência será verificada em nossos dados.

A partir das informações apresentadas nesta subseção sobre os plurais em *–ão*, percebemos que: 1) A controvérsia nesse grupo de plurais já ocorria no português arcaico, visto que na obra de escritores do século XV são encontrados exemplos de itens grafados com plural diferente de sua origem etimológica; 2) Em seu percurso histórico, a língua portuguesa recebeu maior quantidade de itens pluralizados em *–ões* do que em *–ãos* e *–ães*, por isso existe, até hoje, na nossa língua, maior número de palavras que fazem plural em *–ões*; 3) A forma *–ões* é a preferida pelos falantes quando palavras novas terminadas em *–ão* entram na língua e são pluralizadas, o que demonstra a produtividade do morfema *–ões*; 4) Através da análise das gramáticas de língua portuguesa, observamos a tendência de palavras pluralizadas em *–ãos* e *–ães* adotarem plural em *–ões*. Os únicos exemplos de itens com etimologia *–ões* que adotaram plural em *–ãos* ou *–ães* foram "peão" e "pião", conforme comentamos anteriormente. Tais fatos apontam evidências diacrônicas de que, na classe de palavras terminadas em *–ão* no singular, o morfema *–ões* de plural é o mais produtivo e freqüente. É interessante notar que a regra geral de pluralização do PB é o acréscimo do morfema *–s*, portanto poderíamos supor que seria mais plausível que a terminação *–ãos* (*–ão* + *–s*) fosse preferida em detrimento dos morfemas *–ães* e *–ões*. Já que essas duas últimas formas requerem o acréscimo do morfema *–s* e transformações no radical da palavra, o falante poderia simplesmente acrescentar um morfema *–s* às palavras terminadas em *–ão* no singular, pluralizando-as, portanto, como *–ãos*, já que o plural em *–s* é a regra mais produtiva da língua. No entanto, os fatos diacrônicos fornecem evidência de que a forma em *–ões* é, historicamente, a mais proeminente e que palavras em *–ãos* e *–ães*, em alguns casos, migraram para a classe de *–ões*. Tais fatos desafiam a hipótese de que o falante preferiria aplicar a regra canônica de pluralização em *–s* para todos os casos de itens terminados em vogais no PB. No capítulo seguinte, acrescentaremos informações sincrônicas a esses fatos diacrônicos e poderemos traçar hipóteses mais precisas sobre as diferentes desinências de pluralização para as palavras em *–ão* no singular.

---

[18] A palavra "bênção" tem sua etimologia em *benedictio*, *onis* (plural em *–ões*), no entanto sua grafia, desde o século XV, já era *bencam* e, como é paroxítona, sua pluralização considerada correta é em *–ãos*.

Nas próximas subseções, analisaremos os diversos aspectos relacionados às terminações *–l* e ditongo em *–u* do PB e seus respectivos plurais.

2.3 Plural das palavras terminadas em *–l* e ditongo em *–u*[19]

Conforme podemos verificar nos exemplos abaixo, o PB possui um grupo de palavras terminadas em *–al*, *–el*, *–il*, *–ol*, *–ul* (ortográficos):

normal	[nofiˈmaw]	sol	[ˈsɔw]
anel	[aˈnɛw]	azul	[aˈzuw]
funil	[fuˈniw]		

A pluralização de tais palavras obedece às seguintes regras:

1 – Se a palavra terminar em *–al*, *–ol* ou *–ul*, sua terminação muda para *–ais*, *–óis* e *–uis*, respectivamente:

sal	→	sais	[ˈsays]	animal	→	animais	[aniˈmays]
lençol	→	lençóis	[lẽˈsɔys]	anzol	→	anzóis	[ãˈzɔys]
azul	→	azuis	[aˈzuys]	paul[20]	→	pauis	[paˈuys]

2 – Se a palavra terminar em *–el*, faz seu plural em *–éis* (quando *–el* é tônico) ou *–eis* (quando *–el* é átono):

[19] A quantidade de informações encontradas para os plurais terminados em *–l* e ditongo em *–u* não foi extensa e os dois plurais estão relacionados (conforme veremos a seguir), por isso decidimos mencionar em uma única subseção as informações sobre esses dois grupos.
[20] Segundo o Novo Dicionário Aurélio Eletrônico, "paul" é uma variação para "pântano".

papel	→	papéis	[paˈpɛys]	anel	→	anéis	[aˈnɛys]
móvel	→	móveis	[ˈmɔveys]	túnel	→	túneis	[ˈtuneys]

3 – Se a palavra terminar em *–il* tônico, perde o *–l* e adota *–s*; se terminar em *–il* átono, muda *–il* para *–eis*:

funil	→	funis	[fuˈnis]	sutil	→	sutis	[suˈtʃis]
fácil	→	fáceis	[ˈfaseys]	útil	→	úteis	[ˈuteys]

O PB apresenta, também, um grupo de palavras terminadas em *–au*, *–eu*, *–iu*, *–ou*, *–uu*[21], conforme mostram os exemplos abaixo:

bacalhau	[bakaˈʎaw]	abiu	[aˈbiw]
céu	[ˈsɛw]	grou[22]	[ˈgrow]

A pluralização de tais itens obedece à regra canônica do PB, ou seja, basta que se acrescente um morfema *–s* ao fim das palavras:

bacalhau	→	bacalhaus	[bakaˈʎaws]	abiu	→	abius	[aˈbiws]
céu	→	céus	[ˈsɛws]	grou	→	grous	[ˈgɾows]

Com relação às palavras terminadas em *–l* ou ditongo em *–u*, é fundamental para esta análise mencionar que, em fim de sílaba, o segmento /l/, no PB, é pronunciado como [w]:

---

[21] No Novo Dicionário Aurélio Eletrônico, encontramos as seguintes palavras terminadas em *–uu*: "anuu", "inambuu", "inamuu", "inhambuu", "nambuu", "nhambuu" e "Uuu". Maiores discussões sobre esse pequeno grupo de itens serão feitas no Capítulo 3, "Revisão de Literatura".

[22] Segundo o Novo Dicionário Aurélio Eletrônico, "abiu" é o "fruto do abieiro, de baga branca e doce", e "grou" é uma "ave pernalta, cultrirrostra (*Grus cinerea*)".

> Na maioria dos dialetos do português brasileiro, o que ocorre é um processo de vocalização do l. De acordo com tal processo, articulamos um segmento com a qualidade vocálica de u na posição correspondente ao "l" ortográfico em posição final de sílaba: "sal, salta". Adotamos o símbolo [w] para transcrever tal segmento. Neste caso formas como "sal, salta" são transcritas foneticamente como ['saw] e ['sawta], respectivamente. (CRISTÓFARO-SILVA, 2002, p. 63)

Assim, devido à perda de contraste fonético entre *–l* e *–u* em fim de sílaba, os pares de itens abaixo[23] são pronunciados com a mesma terminação:

degrau	[de'ɡɾaw]	abiu	[a'biw]
legal	[le'ɡaw]	canil	[ka'niw]
chapéu	[ʃa'pɛw]	grou	['ɡɾow]
mel	['mɛw]	gol	['ɡow]

A vocalização de *–l* em fim de sílaba é uma tendência que ocorreu, também, na passagem do latim ao português, como nos exemplos alteru > outro, falce > fouce e palpare > poupar. Abaurre (1973, p. 19) apresenta a seguinte explicação para esse fenômeno:

> In Brazilian Portuguese there is a synchronic rule, for most dialects, that vocalizes l and change it into w in syllable-final position. Only very conservative dialects maintain the l in this position, and it may well be the case that we are observing what in the long run will be considered to be a historical phonological change leading to the simplification of this kind of syllable in the language, since a syllable closed by a glide is closer to the optimal CV structure than a syllable closed by a liquid.

Dessa forma, a justificativa de Abaurre (1973) para a vocalização de *–l* em fim de sílaba reside na preferência do PB por sílabas formadas por consoante e vogal. Segundo a autora, uma sílaba formada por consoante, vogal e glide (com a vocalização de *–l*) estaria mais próxima dessa estrutura canônica CV.

O segmento /l/, embora seja realizado como [w] em fim de sílaba, tem sua qualidade consonantal preservada em palavras derivadas, como podemos perceber nos exemplos a seguir:

---

[23] Não citamos exemplos para o par *–uu*/*–ul* porque a ocorrência de itens terminados em *–uu* no PB é muito baixa, conforme mencionaremos no Capítulo 3, "Revisão de Literatura".

farol	→	faroLete	papel	→	papeLaria
mel	→	meLado	sal	→	saLeiro

Por sua vez, as palavras que terminam em ditongo em *–u* sofrem outro processo de derivação, pois não apresentam *–l* no radical:

museu	→	museÓlogo	pneu	→	pneUmático

Através das derivações apresentadas acima, podemos perceber que a perda de distinção entre *–l* e *–u* em fim de sílaba é apenas fonética, porque, em termos morfológicos, pelo menos no que concerne à derivação, a distinção permanece[24].

Devido a essa falta de distinção fonética entre *–l* e *–u* em fim de sílaba, a pluralização das palavras terminadas em *–l* e ditongo em *–u* às vezes apresenta divergências. Se não há contraste fonético entre *–l* e *–u*, a informação possível para adotar o plural adequado seria o correlato ortográfico: se a palavra termina em *–l*, a regra básica seria a eliminação do /l/ e o acréscimo de *–is*; se a palavra termina em ditongo em *–u*, basta que se acrescente *–s*. No entanto, em geral os falantes parecem não ter consciência clara dessa regra e, por isso, alguns plurais desviantes são ocasionalmente adotados, como "chapéis" em vez de "chapéus", "degrais" em vez de "degraus" e "troféis" em vez de "troféus", para citar apenas alguns casos.

Dessa maneira, observa-se que, nesse grupo específico de palavras terminadas em ditongo em *–u*, os falantes, algumas vezes, tendem a adotar a regra de pluralização em *–is*, típica de palavras terminadas em *–l*. É interessante mencionar que, em geral, palavras terminadas em *–l* não costumam adotar a regra de pluralização das palavras terminadas em *–u*. Uma possível explicação para esse fato será apresentada no capítulo seguinte.

Na próxima subseção, analisaremos a etimologia das palavras terminadas em *–l* e ditongo em *–u*, a fim de traçarmos um percurso evolutivo de tais morfemas desde o latim até o português.

---

[24] O item "chapéu" parece contradizer essa afirmação, já que termina em ditongo em *–u* e a derivação gera "chapelaria" (com recuperação do *–l*). Segundo o Novo Dicionário Aurélio Eletrônico, essa palavra deriva do francês antigo *chapel* e, por isso, suas formas derivadas são em *–l*, embora sua ortografia no PB atual seja em ditongo em *–u*.

2.3.1 A etimologia das palavras terminadas em *–l* e ditongo em *–u*

Segundo Coutinho (1981, p. 108-109), no latim havia apenas quatro ditongos: *–ae–*, *–oe–*, *–eu–* e *–au–*. Com a evolução para o português, alguns desses ditongos apresentaram tendência a reduzirem-se para vogais simples: "aequale" > "igual", "caelu" > "céu", enquanto o ditongo *–au*, a partir do século X, algumas vezes se transforma em *–ou*: "thesauru" > "tesouro".

De acordo com o mesmo autor, o português apresenta mais ditongos que o latim por causa de vários fenômenos fonológicos distintos que concorreram para que ocorresse ditongação em alguns grupos consonantais ou vocálicos. Um processo que merece destaque aqui é a síncope de fonemas mediais, que promoveu o aparecimento de alguns novos ditongos no português: "malu" > "mau", "palu" > "pau", "lege" > "lei".

Com relação à etimologia do segmento *–l* em fim de palavra, Coutinho (1981, p. 237-238) afirma que as terminações do acusativo latino *–ales*, *–eles*, *–īles*, *–ĭles*, *–oles*, *–ules*, após a queda do *–l–* intervocálico e a passagem do *–e–* átono para *–i–*, evoluíram para os plurais *–ais*, *–éis*, *–is*, *–eis*, *–óis*, *–uis* do português moderno. Os exemplos mencionados para esses casos são:

> *–ales > –aes > –ais: *animales por animalia > animaes > animais;*
> *–eles > –ees > –éis: fideles > fiees > fiéis;*
> *–īles > –ies > –iis: *cubiles > por cubilia > covies > coviis > covis;*
> *–ĭles > –ees > –eis: fossiles > fossees > fósseis;*
> *–oles > –oes > –óis: soles > soes > sóis;*
> *–ules > –ues > –uis: padules por paludes > paues > pauis.*
> (Cf. COUTINHO, 1981, p. 238)

Guerios (1942, p. 117) apresenta exemplos em que o *–l–* intervocálico se manteve na forma plural:

> Exemplos do port. arc. em que, no plural, conservava-se o *–l–* apesar de intervocálico: *anafiles*, *curules*, *dóciles*, *ediles*, *estábiles*, *estériles*, *fértiles*, *froles* (sing. *frol*, flor), *hábiles*, *inútiles*, *papeles*, *pênsiles*, *reales*, *roles*, *soles*, *terríbiles*, etc. Provavelmente o *–l–* aí permaneceu ou por influencia (*sic*) erudita, ou espanhola. O *–l–*, entretanto, na atualidade resistiu em *cales* (cano) *cônsules* (por influencia (*sic*) erudita), *males* (evita confusão com o adverbio (*sic*) *mais*), *meles* (ao lado de *méis*).

Assim, podemos afirmar que a forma atual das palavras terminadas em *–l* no PB advém de dois fenômenos distintos:

(i) No plural, ocorreu a queda do *–l–* intervocálico e a passagem do *–e–* átono a *–i–*, produzindo percursos evolutivos como:

capitales	>	capitaes	>	capitais
aneles	>	anees	>	anéis
ardiles	>	ardies	>	ardis
fossiles	>	fossees	>	fósseis
anzoles	>	anzoes	>	anzóis
azules	>	azues	>	azuis

(Cf. PARREIRA; PINTO, 1985, p. 162)

(ii) No singular, o *–l–* tornou-se final e se manteve, após a queda da vogal átona, como nos seguintes casos: "crudele" > "cruel", "fidele" > "fiel", "sole" > "sol".

Camara Júnior (1972, p. 66-67) explica essa evolução e o subseqüente desenvolvimento de regras diferentes de pluralização para as palavras terminadas em *–l* no PB:

> The dropping of intervocalic /l/, a phenomenon typical of the evolution of the Portuguese consonant system, caused the thematic vowel *–e* to come into contact with the preceding stressed vowel. This development, in turn, eventually resulted in several distinct types of evolution.

Nunes (1945, p. 229) apresenta uma explicação etimológica para o grupo de palavras terminadas em *–l*:

> Como é sabido, o plural em *–aes*, *–eis* (=*ees*), *–is* (=*iis* ou *ies*), *–oes* e *–ues* dos nomes terminados em *–al*, *–el*, *–il*, *–ol*, *–ul* é devido à queda do *l* intervocálico; a terminação *–iles* deu, quando tónica *–ies*, que passou para *–iis* e depois se reduziu a *–is*, e, quando átona, *–ees*, que mais tarde, por dissimilação ou devido ao lugar ocupado pelo último *e*, se tornou em *–eis*, evolução que igualmente sofreu a tónica *–eles*.

O autor ainda faz dois acréscimos: em primeiro lugar, a ortografia atual optou por grafar sempre por *–i–* aquele *–e–* que forma ditongos orais junto a *–a–* ou *–u–*, como nos casos de "fatais" e "azuis" (em vez de "fataes" e "azues"); em segundo lugar, a distinção de plural por tonicidade não existia no português arcaico, já que nomes oxítonos e paroxítonos faziam plural da mesma maneira: "movis", "razoavis", "estavees", "notavees", em vez de "móveis", "razoáveis", "estáveis", "notáveis".

A partir dessas informações etimológicas, observamos que a terminação ditongo em *–u* em fim de palavra é menos freqüente no PB que a terminação *–l* (na mesma

posição) porque muitas palavras latinas que originalmente possuíam ditongos perderam a ditongação e adotaram vogais simples. A prevalência dos ditongos *–au* e *–eu* em fim de palavra pode ser explicada pelo fato de que esses já eram ditongos no latim, enquanto que *–iu*, *–ou* e *–uu* não eram encontrados no latim e se desenvolveram posteriormente, com a evolução da língua. Apesar do surgimento de novos ditongos a partir da queda dos fonemas mediais (conforme exemplos acima), os ditongos formados a partir de vogal + *–u* não foram em número suficientemente alto a ponto de ultrapassarem, em termos de classe, a terminação *–l* em fim de palavra[25].

Na subseção seguinte, analisamos a morfologia dos substantivos e adjetivos terminados em *–l* e ditongo em *–u*.

2.3.2 A morfologia das terminações *–l* e ditongo em *–u*

As palavras terminadas em ditongo em *–u* apresentam uma morfologia bastante simplificada, porque há apenas dois sufixos, ambos representados pela terminação *–eu*. Segundo Cunha e Cintra (1985, p. 98), o sufixo *–eu* ([ew]) pode indicar relação, procedência ou origem, como nos casos "europeu" e "hebreu". Essa terminação aparece, também, no sufixo *–aréu* ([arɛw]), formando aumentativos como "fogaréu" e "povaréu". Com exceção desses dois sufixos formados com o morfema *–eu*, todas as demais terminações (*–au*, *–iu*, *–ou*, *–uu*) fazem parte do radical da palavra, ou seja, não representam sufixos na língua portuguesa.

Por sua vez, as terminações *–al*, *–el*, *–il*, *–ol* e *ul* podem fazer parte do radical da palavra, como no caso de "anel", "azul", assim como podem, também, representar diferentes sufixos. Coutinho (1981) e Cunha e Cintra (1985) apresentam uma lista com oito sufixos terminados em *–l* e sua origem latina. Exemplos de tais sufixos são: 1) *–al*, que pode indicar relação ou coletivo: "areal", "cafezal", "vital"; 2) *–vel*, que exprime possibilidade de praticar ou sofrer uma ação: "admirável", "amável", "incrível", "perecível"; 3) *–il*, que designa diminutivos ou local onde se guardam animais: "canil", "covil", "pernil".

---

[25] No capítulo seguinte ("Revisão de Literatura"), apresentaremos estatísticas precisas sobre a quantidade de itens do PB que apresenta *–l* ou ditongo em *–u* em fim de palavra.

Nesta subseção sobre o status morfológico das terminações *–l* e ditongo em *–u*, percebemos que a morfologia da terminação *–l* é mais complexa, porque pode carregar vários tipos de sufixos; por outro lado, no que concerne à terminação de ditongo em *–u*, há somente dois sufixos ([ew] ou [aɾɛw]) e, mesmo assim, ambos são pouco produtivos na língua, ou seja, uma quantidade pequena de palavras sofre flexão através dele (Cf. Capítulo 3, "Revisão de Literatura"). Na próxima subseção, fazemos uma conclusão acerca das informações apresentadas neste capítulo.

2.4 Conclusões

Neste capítulo, comentamos informações diacrônicas e morfológicas referentes aos grupos de plurais em *–ão*, *–l* e ditongo em *–u*. Pudemos analisar as peculiaridades inerentes a cada um desses grupos e, com base nessas informações, algumas conclusões preliminares são possíveis:

- Apesar de a gramática normativa do PB apresentar regras de pluralização para cada um dos grupos de plural, a língua registra casos, desde o português arcaico, em que não havia conformidade à regra;
- Mesmo entre gramáticas normativas, o tema plurais em *–ão* do PB deixa dúvidas quanto à forma correta a ser adotada;
- Existem evidências diacrônicas de que a forma *–ões* é a mais produtiva para itens terminados em *–ão* no singular, por isso palavras etimológicas em *–ãos* e *–ães* estão adotando plurais em *–ões*;
- Informações morfológicas demonstram que a forma plural em *–ões* (para *–ão* no singular) e as formas *–is* ou *–eis* (para *–l* no singular) representam mais sufixos que os itens em *–ãos* e *–ães* (para singular em *–ão*) e *–us* (para singular em ditongo em *–u*).
- A etimologia dos morfemas *–l* e ditongo em *–u* demonstra que a terminação *–l* é, desde o português arcaico, mais freqüente que ditongo em *–u* (ambas no fim de sílaba).

Portanto, este capítulo apresentou aspectos morfológicos, etimológicos e fonológicos da flexão nominal de grupos de plurais do PB a partir da perspectiva da gramática normativa. No entanto, em consonância com Bybee (2001) e Hopper (1998), supomos que a gramática emerge do uso que os falantes fazem dela em suas interações cotidianas; por conta disso, apresentaremos, a seguir, mais um capítulo com informações sobre os plurais do PB sob análise nesta tese. Nesse próximo capítulo, comentaremos algumas pesquisas lingüísticas que, a partir de análises do uso da língua falada, apontam fenômenos que estão ocorrendo na concordância nominal de número do PB. Essas investigações são importantes, a fim de que possamos traçar um correlato entre as informações diacrônicas (apresentadas neste capítulo) e processos que possam estar ocorrendo sincronicamente com a flexão de plural do PB, mais especificamente com os grupos de plurais sob análise nesta tese. Passemos, portanto, às informações do próximo capítulo.

CAPÍTULO 3

Revisão de Literatura

Conforme já foi mencionado anteriormente, os Capítulos 2 e 3 desta tese têm como objetivo principal oferecer informações sobre a concordância nominal de número para os plurais definidos como objeto de estudo na presente análise. O capítulo anterior expôs aspectos diacrônicos e morfológicos sobre os grupos de plurais do PB, considerando como fontes, principalmente, gramáticas históricas e normativas do português. O presente capítulo tem como objetivo proceder a uma resenha sobre as principais pesquisas lingüísticas já realizadas sobre a concordância nominal de número e os grupos de plurais do PB sob investigação nesta tese. Além disso, ao final desta resenha, apresentamos uma análise crítica das pesquisas feitas sobre esse tema no PB, apontando algumas lacunas que os referidos trabalhos deixaram e esclarecendo as possíveis contribuições e inovações que a análise proposta por esta tese pode fornecer para o entendimento dos grupos de plurais em *–ão*, *–l* e ditongo em *–u*. A combinação entre as informações diacrônicas, morfológicas e etimológicas (apresentadas no capítulo anterior) e as análises lingüísticas já realizadas sobre o tema (apresentadas neste capítulo) propiciará a formulação de hipóteses preliminares para os grupos de plurais sob análise nesta tese. Ademais, nas últimas subseções deste capítulo, apresentamos algumas informações sobre possíveis interações entre processos morfofonológicos de flexão nominal de número no PB e representações mentais das informações lingüísticas.

Este capítulo, portanto, apresenta a seguinte organização: na subseção 3.1, comentamos análises lingüísticas anteriores já realizadas sobre a pluralização no PB; na subseção 3.2, analisamos criticamente as pesquisas sobre plurais mencionadas na subseção 3.1; na subseção 3.3, mencionamos como a freqüência (de tipo ou de ocorrência) pode interferir nos grupos de plurais analisados nesta tese; na subseção 3.4, enumeramos as principais conclusões alcançadas com este capítulo.

3.1 Análises anteriores sobre processos de pluralização do português brasileiro

Thomas (1969, p. 8-10) comenta alguns casos específicos de concordância nominal peculiares ao PB. Vamos citar os dois principais pontos levantados por esse autor:

1. **O uso do singular pelo plural.** Segundo o autor, em alguns casos, uma palavra no singular é utilizada, mas a idéia é de plural, como nos seguintes exemplos: "Todos os homens deixaram **o carro** em casa.", "**A cinza** do cigarro caiu no chão." (THOMAS, 1969, p. 8-9);
2. **Perda da flexão de plural.** "In substandard Portuguese, there is a widespread tendency to indicate number only with the article or other determinative, leaving all words following – the noun and adjectives – in the singular." Exemplos citados para esse caso são: "Lá tem **umas brincadeira interessante**.", "Ele tinha **umas idéia meio comunista**." (THOMAS, 1969, p. 10). O autor comenta, também, que "While persons of even moderate education avoid this, it is found among the uneducated in all parts of the country." (THOMAS, 1969, p. 10)

Amaral (1920), em sua análise sobre o dialeto caipira, comenta a ausência da marca de plural, salientando que, em geral, o plural é mantido no determinativo: "os rei mago", "umas fruita verde" (Cf. AMARAL, 1920, p. 52).

Jucá-Filho (1937, p. 23), fazendo uma análise da "Língua Nacional", comenta o seguinte sobre a perda da marca de plural no PB: "O caipira esquece freqüentemente a flexão genérica do Adjetivo [...]. Quanto à flexão numérica, essa é em geral sacrificada em todo o Brasil, salvo nos Determinativos."

Marroquim (1945, p. 111-113), em seu livro "A língua do Nordeste", também fornece informações sobre a flexão nominal nesse dialeto. Segundo o autor, no dialeto nordestino:

1. Apenas o determinativo carrega flexão de número: "os home", "os ingrei" ("ingleses"), "os patrão", "as vêi" ("vezes"), "vinte mi rei" ("réis"), "dois mi rei".

Esse processo de redução das flexões de número é considerado, pelo autor, uma simplificação;

2. A perda da concordância nominal é "o fenômeno mais pessoal e frisante do dialeto popular." (MARROQUIM, 1945, p. 112);
3. O *–s* de plural "só permanece nos determinativos por uma necessidade psicológica, porque sem isso nada ficaria para indicar essa pluralidade." (MARROQUIM, 1945, p. 113). Segundo ele, não fosse essa obrigatoriedade de indicar o plural, o dialeto falado no Nordeste não teria palavras terminadas em coda, por conta das reduções fonológicas e morfológicas.

A partir das informações apresentadas nos parágrafos anteriores, percebe-se que o fenômeno de perda de plural é recorrente no PB e já era registrado em nosso país no meio do século passado, a despeito dos comentários preconceituosos dos autores que realizaram as investigações mencionadas acima; por outro lado, observa-se, também, que, em geral, a ausência da marca de plural está relacionada a indivíduos pertencentes a classes sociais desfavorecidas, como o "caipira", o povo nordestino ou pessoas de classe baixa e com pouca escolaridade. Uma vez que elementos sociais podem interferir na marcação dos plurais do PB, em nossa pesquisa incluiremos, também, variáveis extralingüísticas, a fim de investigar se existe algum correlato entre flexão nominal de número e peculiaridades sociais, tais como faixa etária e nível de escolaridade.

Embora o foco de investigação de nossa tese não seja, especificamente, a análise da perda das marcas formais de plurais do PB, não podemos deixar de considerar esse aspecto em nossa pesquisa, já que algumas análises lingüísticas (conforme mencionamos acima) já registraram tal fenômeno na nossa língua. Assim, em nossa coleta de dados sobre os grupos de plurais do PB, uma das variáveis sob análise será a perda das marcas formais de pluralização, bem como as diferentes possibilidades de marcação do plural para cada um dos processos de pluralização analisados.

Na literatura lingüística sobre o PB, não encontramos muitos trabalhos que tratem especificamente dos grupos de plurais analisados em nossa tese. Existem, entretanto, algumas análises para o fenômeno da concordância nominal de número que incluem, também, a investigação sobre os plurais terminados em *–ão*, *–l* e ditongo em *–u*. Tais pesquisas, embora não abordem particularmente esses casos de plurais, analisam a pluralização do PB como um todo e apresentam, também, algumas observações a respeito

dos grupos de plurais em *–ão*, *–l* e ditongo em *–u*. Como essas classes de plurais se inserem no todo da pluralização do PB e não podem ser dissociadas dos outros fenômenos morfofonológicos de formação de plural[26], comentamos, a seguir, algumas análises já realizadas sobre plurais no PB. Possivelmente, as mais extensas e aprofundadas investigações a esse respeito são as de Guy (1981) e de Scherre (1988). Neste capítulo, tais estudos merecerão subseções separadas[27].

3.1.1 A abordagem de Guy (1981)

Guy (1981) investiga a fala de 20 informantes semi-escolarizados de classe baixa nascidos na cidade do Rio de Janeiro. Através da teoria sociolingüística variacionista (LABOV, 1972), o autor analisa o apagamento do *–s* nominal sob duas perspectivas distintas: 1) Quando o *–s* se encontra em palavra monomorfêmica e não representa marca de plural, como em "menos", "ônibus", "rapaz", "seis"; e 2) Quando o *–s* se encontra em palavras bimorfêmicas e indica pluralidade: "árvores", "cores", "irmãos", "olhos".

Segundo o autor, duas abordagens distintas podem ser feitas para a perda do *–s* final: i) Em primeiro lugar, pode-se afirmar que o apagamento do *–s* corresponde a uma perda fonológica, já que o PB tem prevalência de sílabas CV (consoante + vogal) e, por isso, segmentos consonantais finais, como *–r* e *–s*, costumam ser cancelados; ii) Em segundo lugar, pode-se considerar o apagamento do *–s* final como um processo morfossintático de perda das marcas de plural. A fim de investigar a validade dessas abordagens, o autor analisou os dois grupos de palavras isoladamente (palavras em que o *–s* não é marca de plural e palavras em que o *–s* é marca de plural), a partir das mesmas variáveis lingüísticas (tonicidade e segmento seguinte). Utilizando o programa Varbrul 2, o autor concluiu que, em diferentes dimensões, os resultados para as duas categorias de

---

[26] Nesta tese, como temos o objetivo de fazer uma análise sobre grupos específicos de plurais do PB (em *–ão*, *–l* e ditongo em *–u*), não fizemos menção aos demais processos morfofonológicos de pluralização. No entanto, como a pluralização é uma característica flexional do PB como um todo, informações sobre um processo de pluralização podem colaborar para o entendimento de outro. Em função disso, a análise de trabalhos que tratam da pluralização em termos mais gerais pode contribuir para nossa pesquisa.

[27] Não apresentaremos com minúcias todos os resultados que as abordagens anteriores sobre a concordância nominal de número alcançaram. Mencionaremos apenas os resultados gerais dessas pesquisas e os aspectos relevantes para a proposta que esta tese apresenta, que é a análise dos grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u*.

palavras quanto aos fatores codificados eram os mesmos. Em linhas gerais, ele observou que, nos dois casos de *–s* analisados, sílabas átonas favorecem o apagamento do *–s* e, com relação ao segmento seguinte, a marca de *–s* plural se conserva mais diante de pausa, ao passo que o *–s* monomorfêmico é mais preservado diante de vogal.

Para a análise dos casos de *–s* como marca de plural, as duas variáveis mais importantes analisadas pelo autor foram a posição da palavra no sintagma nominal e o processo de formação de plural (GUY, 1981, p. 161). Comentaremos cada uma delas em separado.

Quanto à posição da palavra no sintagma nominal, Guy (1981, p. 168) observou uma escala crescente de apagamento do plural do primeiro ao último elemento. Para citar um exemplo, em um sintagma como "Meus melhores amigos brasileiros", o primeiro elemento ("meus") seria pluralizado em 95% dos casos, o segundo elemento ("melhores"), em 28%, o terceiro ("amigos"), em 22%, e o quarto ("brasileiros"), em 13%. Dessa forma, o plural seria marcado com mais freqüência na primeira palavra do sintagma nominal, e os demais itens iriam perdendo a pluralização numa escala crescente.

Com relação à morfologia de pluralização, Guy (1981, p. 181-182) esclarece que dividiu os processos morfofonológicos de formação de plural em sete categorias distintas, a saber:

1. Plurais regulares, formados a partir do acréscimo de *–s*, que representam mais de 9/10 dos dados coletados[28];
2. Palavras terminadas em *–ão* que fazem plural em *–ães* ou *–ões*;
3. Plurais de palavras terminadas em /R/;
4. Plurais de palavras terminadas em /S/ (*–s* ou *–z* ortográficos);
5. Palavra "vez", que foi separada dos itens acima por causa de sua alta freqüência na expressão "às vezes";
6. Plurais das palavras terminadas em *–l*, que apresentam processos distintos de pluralização, dependendo da tonicidade;
7. Plurais em metafonia.

[28] Neste grupo, foram incluídos os itens terminados em *–ão* que fazem plural com acréscimo de *–s*, como "irmãos" e "mãos" (etimologia *–ano*).

Considerando-se essas categorias, os resultados alcançados com a análise de dados (Cf. GUY, 1981, p. 184) são apresentados na tabela abaixo:

Tabela 1: Processos morfofonológicos de formação de plural e preservação do plural

Category of Plural Formation	% Plural Marked	N	Prob.
Regular	65	9058	.28
–ão	52	84	.30
–l	44	57	.33
–r	68	142	.62
–z	52	147	.51
vez	48	298	.54
Metaphonic	75	8	.86

Na análise dessa tabela, Guy (1981) enfatiza que: 1) Os itens regulares e os terminados em *–ão* e *–l* apresentam diferenças probabilísticas mínimas (0.28 – 0.30 e 0.33, respectivamente); 2) Os itens terminados em *–r*, *–z* e a palavra "vez" também apresentam probabilidades bastante próximas entre si: 0.62 – 0.51 e 0.54, respectivamente; 3) Os itens em metafonia apresentam quantidade pequena de dados (oito ocorrências) para que alguma conclusão possa ser traçada.

A partir dessas observações, Guy (1981, p. 186) decide excluir os plurais em metafonia, por causa do pequeno quantitativo de dados, e reagrupar os demais tipos de plural em duas grandes classes, uma formada por palavras terminadas em vogal e outra formada por palavras terminadas em consoante. Os resultados desse novo agrupamento demonstram que itens terminados em consoante favorecem a preservação do plural (0.64) e itens terminados em vogal desfavorecem a presença de marcas formais de plural (0.36). Esse é, para o autor, o melhor ajuste que os dados podem apresentar e, portanto, a grande diferença entre marcas de plural do PB ocorre em função do segmento final (vogal ou consoante).

Quanto aos fatores sociais analisados por Guy (1981), os resultados indicam que: 1) As mulheres favorecem ligeiramente a presença da concordância (0.56) e os homens a desfavorecem (0.44); 2) Com relação à faixa etária, não houve nenhuma distribuição sistemática, visto que as probabilidades foram muito próximas para todos os grupos de informantes analisados.

3.1.2 A abordagem de Scherre (1988)

Scherre (1988) utiliza dados de entrevistas realizadas com 64 informantes adultos e crianças da cidade do Rio de Janeiro, selecionados a partir dos critérios escolaridade, gênero e faixa etária. A autora propõe uma análise da concordância nominal de número no PB também a partir da sociolingüística variacionista. Em Scherre (1988), a concordância nominal de número é considerada a partir de duas óticas diferentes: uma análise atomística, considerando as palavras do sintagma nominal isoladamente, e uma análise não-atomística, focalizando o sintagma nominal como um todo e observando as interações entre os seus constituintes. Como o foco principal de análise da nossa tese é processos morfofonológicos de pluralização no PB (especificamente nos grupos em *–ão*, *–l* e ditongo em *–u*), concentraremos nossa atenção na análise atomística apresentada pela autora.

Diferentemente de Guy (1981), a autora considerou apenas dados em que o *–s* final era marca de pluralização. Casos de *–s* final em palavras monomorfêmicas (como ônibus e menos) não foram analisados porque o objeto de estudo de sua tese era a concordância nominal de número.

Para a codificação dos dados, a autora considerou "todo e qualquer SN [sintagma nominal] que [tivesse] pelo menos uma marca formal ou semântica de plural" (SCHERRE, 1988, p. 29). A partir desse critério básico, um total de 7.193 sintagmas nominais foram codificados, compondo 13.229 dados individuais (cada uma das palavras que formavam o sintagma nominal).

Em sua análise, a autora discute a "escala de saliência fônica" proposta por Braga e Scherre (1976) para os diferentes grupos de plurais do PB. Por "saliência fônica" devemos entender o grau de diferenciação de material fônico na relação singular/plural, ou seja, a diferença fônica que uma palavra no plural apresenta em relação ao seu singular. Reproduzimos abaixo a escala apresentada por Braga e Scherre, 1976 *apud* Scherre, 1988, p. 66-67:

1) inserção de –S e abertura vocálica ou plural metafônico (ovo/óvos);
2) inserção de –S e mudança silábica (milhão/milhões; imóvel/imóveis);
3) inserção de –ES em palavras em –R (flor/flores);
4) inserção de –S em palavras de plural regular (livro/livros);
5) inserção de –ES em palavras que terminam em –S (rapaz/rapazes).

Na escala acima, percebemos que a saliência fônica diminui do item 1 ao 5. O grupo 1 apresenta a maior saliência fônica, porque pressupõe duas modificações entre o singular e o plural: acréscimo de *–s* e abertura da vogal /o/, portanto a diferenciação entre o singular e o plural (saliência fônica) é maior. Por outro lado, o item 5 apresenta a menor saliência, porque a diferença fônica entre o singular terminado em /S/ e o plural *–es* é mais discreta do que nos outros casos. Por causa dessa distinção entre material fônico no singular e no plural, esperar-se-ia que os itens do grupo 1 acima (com metafonia e acréscimo de *–s*, simultaneamente) preservassem mais a concordância do que os itens em 5 (com acréscimo de *–es*). Isso seria explicado pelo fato de que a diferença entre o singular e o plural do grupo 1 é foneticamente muito significativa para ser perdida, ao passo que, na classe 5, a diferença fônica é bem menor que na 1 e, por isso, mais facilmente perdida. Assim, segundo Braga e Scherre (1976 *apud* SCHERRE, 1988), os plurais de maior saliência fônica favorecem a presença da concordância nominal de número, enquanto que os de menor saliência fônica desfavorecem. Na escala acima, é importante observar que Braga e Scherre (1976 *apud* SCHERRE, 1988) agruparam em um só nível os plurais das palavras terminadas em *–l* e *–ão*, por considerarem que a diferença entre o singular e o plural desses grupos representa uma saliência fônica equivalente (inserção de *–s* e mudança silábica). Os plurais em ditongo em *–u* encontram-se no grupo 4, de pluralização regular (feita com o acréscimo do morfema *–s*).

Quanto à codificação de dados feita por Scherre (1988), algumas opções metodológicas foram adotadas:

1. Os itens em *–ão* que apresentam plural em *–ãos* (com acréscimo de *–s*, como "irmãos", "mãos") foram, primeiramente, considerados como parte do grupo 1, de palavras que apresentam plural regular. No desdobramento da análise, esses itens foram codificados em separado;
2. Foram alocados no grupo 2 (inserção de *–s* e mudança silábica) apenas os itens terminados em *–ão* que apresentavam plural em *–ões* ou *–ães* ("leões", "pães") por esses envolverem, a um só tempo, acréscimo de *–s* e alterações silábicas;
3. Os plurais em *–ão* e em *–l*, apesar de agrupados em conjunto na escala de saliência fônica mencionada acima, foram codificados em separado pela autora. Por causa disso, os dados de Scherre (1988) foram classificados em seis grupos, em vez de cinco;

4. A autora considerou como plural qualquer tipo de marca de número, fosse ela padrão ou não-padrão. Assim, itens como "comerciai" ("comerciais"), "religiõe" ("religiões") e "pescadore" ("pescadores") foram codificados como plurais, apesar de as marcas de número que eles apresentam não serem padrão.

A análise estatística adotada por Scherre (1988) foi feita a partir de um conjunto de programas computacionais (SANKOFF, 1975, 1979; ROUSSEAU; SANKOFF, 1978; NARO; VOTRE, 1980 *apud* SCHERRE, 1988). Apresentamos, a seguir, os principais resultados alcançados pela autora.

As porcentagens gerais de presença de concordância nominal de número que Scherre (1988) encontrou na análise atomística (considerando cada item léxico do sintagma nominal) foram 72% para os adultos e 66% para as crianças.

Os resultados individuais para cada grupo de plurais são apresentados na tabela a seguir (Cf. SCHERRE, 1988, p. 79):

Tabela 2: Processos morfofonológicos de formação do plural de todos os dados dos adultos

PROCESSOS	Apl./Total		%	Prob.
Duplo (maravilhoso) (maravilhósos)	63/68	=	93	0,86
–l (internacional) (internacionais)	83/97	=	86	0,56
–ão (contradição) (contradições)	173/200	=	86	0,42
–R (pescador) (pescadores)	237/268	=	88	0,48
–S (português) (portugueses)	219/265	=	83	0,38
Regular (pequenininha) (pequenininhas)	7086/10071	=	70	0,24

Na tabela acima, observamos que as probabilidades de presença da flexão de número, de uma forma geral, corroboram a escala de saliência fônica apresentada anteriormente, já que os plurais em metafonia favorecem a presença das marcas de plural (0,86) e os plurais regulares desfavorecem a presença das marcas de plural (0,24). No entanto, duas exceções devem ser mencionadas: i) Os plurais das palavras em *–ão*, que, conforme previsto pela escala fônica, deveriam apresentar probabilidade de concordância semelhante à dos itens em *–l*, o que não ocorre; ii) Os plurais de palavras terminadas em *–s*, que deveriam apresentar taxas mais altas de cancelamento do plural. Nos resultados alcançados por Scherre (1988), a posição mais alta de cancelamento das marcas de plural é ocupada pelos itens regulares. Entretanto, a distinção entre esses dois grupos não é muito significativa, visto que tanto os plurais em *–s* quanto os regulares encontram-se na base da escala de saliência fônica. Excluindo-se esse caso, a única exceção, portanto, seria com relação ao grupo das palavras terminadas em *–ão*, que apresenta comportamento divergente quanto aos itens em *–l*. Isso não seria esperado porque ambos os processos de pluralização ocorrem de forma semelhante, através da inserção do *–s* e mudança silábica. Deveria haver, portanto, probabilidades próximas de presença da concordância formal de número para esses dois grupos, o que não se verifica nos dados.

Além da saliência fônica, fatores como tonicidade, número de sílabas, marcas precedentes, contexto fonético/fonológico seguinte, função sintática do sintagma nominal e aspectos semânticos foram, também, considerados por Scherre (1988). Como nossa tese analisa apenas as palavras isoladamente (e não o sintagma nominal como um todo), comentaremos, dentre os fatores arrolados por Scherre (1988), apenas aqueles que podem ser significativos para a análise aqui proposta.

Quanto à tonicidade, a autora, à semelhança de Guy (1981), observou que as palavras oxítonas e os monossílabos tônicos favorecem a presença do plural (0,66), enquanto os itens paroxítonos e monossílabos átonos (considerados em um só fator), assim como os proparoxítonos, favorecem a perda da concordância (probabilidades de 0,39 e 0,44, respectivamente) (Cf. SCHERRE, 1988, p. 89).

Quanto ao número de sílabas, as probabilidades de preservação da marca de plural foram relativamente próximas: monossílabos (0,48), dissílabos (0,51) e mais de duas sílabas (0,51). Portanto, não houve, em termos estatísticos, interferência significativa do fator número de sílabas na presença ou perda da pluralização (Cf. SCHERRE, 1988, p. 81).

Após a análise dos fatores acima, a autora chega à conclusão de que, em função de o fator tonicidade (entre outros) interferir nos processos morfofonológicos de formação de plural, é necessário elaborar uma nova escala de saliência fônica, que considere, conjuntamente, tonicidade e processos de formação do plural. Segundo essa concepção, a nova escala de saliência fônica apresenta a seguinte organização:

> 1) duplo (oxítono ou paroxítono);
> 2) –l (oxítono ou paroxítono);
> 3) –R (oxítono ou paroxítono);
> 4) –ão (oxítono);
> 5) –S (oxítono);
> 6) regular oxítono;
> 7) regular paroxítono e
> 8) regular proparoxítono. (SCHERRE, 1988, p. 136)

Sendo assim, os dados são recodificados e uma nova tabela é organizada. Os resultados alcançados são reproduzidos abaixo (Cf. SCHERRE, 1988, p. 139):

Tabela 3: Análise final de processos e tonicidade dos substantivos, adjetivos e categorias substantivadas dos adultos por escolarização

SALIÊNCIA FÔNICA		FALANTES			
		Todos	**Primário**	**Ginasial**	**Colegial**
Duplo	**F.**	63= 93%	22= 88%	24= 92%	17=100%
(novo)		68	25	26	17
(nóvos)	**P.**	0,80	0,81	0,82	+ (*sic*)
–l	**F.**	80= 86%	21= 70%	18= 95%	41= 93%
(casal)		93	30	19	44
(casais)	**P.**	0,69	0,64	0,85	0,68
–R	**F.**	237= 88%	77= 82%	74= 89%	85= 94%
(cor)		268	94	83	90
(cores)	**P.**	0,65	0,67	0,64	0,71
–ão	**F.**	173= 86%	31= 72%	61= 84%	81= 96%
(dentão)		200	43	73	84
(dentões)	**P.**	0,59	0,56	0,51	0,77
–S	**F.**	219= 83%	87= 78%	68= 83%	64= 91%
(país)		265	113	82	70
(países)	**P.**	0,56	0,64	0,44	0,51
Regular oxíto-	**F.**	134= 72%	38= 57%	51= 25%	44= 86%
no e monossí-		185	67	66	51
labo tônico	**P.**	0,38	0,38	0,38	0,36
Regular propa-	**F.**	81= 52%	15= 31%	28= 47%	38= 79%
roxítono		156	48	60	48
	P.	0,21	0,16	0,19	0,32
Regular paro-	**F.**	2604= 50%	600= 32%	913= 52%	1076= 67%
xítono		5242	1875	1750	1602
	P.	0,17	0,19	0,16	0,17

Podemos perceber que a nova codificação de acordo com processos morfofonológicos de plural e tonicidade explica melhor o comportamento dos dados coletados por Scherre (1988), já que, nesta nova divisão dos grupos, houve um apagamento gradual das marcas de pluralização de acordo com cada processo morfológico.

Com relação às palavras terminadas em *–ão*, a autora afirma que essa categoria pode ser considerada flutuante, por causa da variabilidade na formação do plural. Admitindo que a existência de duas ou três formas diferentes de plural para a mesma

palavra pode influenciar os resultados, a autora ressalta que, nos 200 dados coletados para esse grupo (em *–ão*), houve apenas um plural em *–ães* ("alemães") e "nenhum caso de plural em *–ão* no sentido de regularização (verão/verãos, por exemplo)" (SCHERRE, 1988, p. 124). A autora afirma que, em todos os dados que já analisou, houve apenas um caso de regularização, que foi "cordão"/"cordãos". Em geral, o mais comum é encontrar a forma totalmente no singular ou totalmente no plural. Nos casos aumentativos, como "amigão", "cachorrão", tais itens apresentaram-se no singular "e não na sua forma plural irregular (amigões, cachorrões, brabões, dentões) e muito menos na forma regular (amigãos, cachorrãos, brabãos, dentãos)." (SCHERRE, 1988, p. 124). A autora, em consonância com Naro (1981), acredita que, na dúvida entre diferentes formas de plural, o falante usa o singular em vez do plural:

> As duas formas plurais cedem a vez à forma singular, apesar da saliência. Esta é uma hipótese bastante sedutora e que poderia se aplicar a todos os casos de ausência de marca nos nomes irregulares em –ão. Podemos imaginar que esteja ocorrendo um processo de regularização, e, uma vez que os itens são regularizados, eles são pouco marcados, como todo e qualquer regular. (SCHERRE, 1988, p. 124)

A fim de testar tal hipótese, Scherre (1988) comparou os dados de itens em *–ão* pluralizados a partir da regra geral (adição do morfema *–s*, como "irmãos", "mãos") com os itens em *–ão* que sofrem mudança de radical, como "leões", "pães" e com as palavras de pluralização regular ("caneta"/"canetas"). Reproduzimos essa comparação na tabela abaixo (Cf. SCHERRE, 1988, p. 125):

Tabela 4: Comparação entre os plurais regulares, regulares em *–ão* e plurais em *–ão* com mudança de radical

SOBRE OS REGULARES EM –ÃO						
PROCESSOS	**F A L A N T E S**					
	Primário			**Ginasial**		
	Apl./Total	**%**	**Prob.**	**Apl./Total**	**%**	**Prob.**
–ão (nação) (nações)	31/43	= 72	0,45	61/73	= 84	0,37
Regular (salário) (salários)	648/1973	= 33	0,35	979/1859	= 53	0,30
Regular em –ão (irmão) (irmãos)	5/17	= 17	0,11	13/17	= 76	0,20

Na tabela acima, observamos que os itens em *–ão* de pluralização regular ("irmãos") apresentam probabilidade de marcação de plural menor que os demais itens em *–ão* (pluralizados em *–ães* ou *–ões*) e até mesmo menor que os regulares como um todo. Esse apagamento das marcas de plural é maior entre os falantes do primário, embora os do ginasial também apresentem altas taxas de perda da pluralização. Os dados apontam, então, para uma regularização dos nomes em *–ão*, já que o apagamento das marcas de plural nessa classe (*–ãos*) é maior que na classe regular propriamente dita; no entanto, a própria autora admite que a quantidade de dados é pequena (34 ocorrências) para que se possa afirmar isso categoricamente.

As principais conclusões que Scherre (1988) alcança com a análise de fatores lingüísticos considerados em sua tese são:

> [...] a concordância ou as marcas de plural ocorrem menos:
> 1) quando há menos saliência fônica na oposição singular/plural;
> 2) quando os itens nucleares ocupam a segunda posição do SN;
> 3) quando os itens não nucleares encontram-se pospostos ao núcleo do SN;
> 4) quando os itens da terceira posição são precedidos por elementos não marcados, formalmente;

> 5) quando o contexto fonológico seguinte é uma vogal ou uma consoante com os traços [+sonoro], [+labial], [+dental] ou [+nasal];
> 6) quando o SN é de função resumitiva;
> 7) quando os itens lexicais são informais, diminutivos/aumentativos e não humanos. (Cf. SCHERRE, 1988, p. 23-24)

Os resultados obtidos por Scherre (1988) para fatores sociais indicam que a preservação das marcas de plural é maior no grupo de falantes mais escolarizados. Isso comprova o fato mencionado na subseção 3.1 deste capítulo, de que a perda de concordância nominal é maior no grupo de falantes menos escolarizados. Com relação à faixa etária e ao gênero, as probabilidades de presença do plural para adultos e crianças são muito próximas, e as mulheres preservam mais a marca de plural que os homens.

Após apresentarmos as linhas gerais dos trabalhos de Guy (1981) e Scherre (1988), passaremos à próxima subseção, que apresenta, em ordem cronológica, informações de outras análises realizadas sobre a flexão nominal de número e grupos de plurais em *–ão*, *–l* e ditongo em *–u* no PB.

3.1.3 Outras abordagens sobre a concordância nominal de número no português brasileiro

Campos (1981) faz uma análise histórica dos plurais em *–ão* do PB. Utilizando dados de outras quatro línguas neolatinas (espanhol, francês, italiano e romeno), a autora compara os diferentes percursos evolutivos que as formas latinas em *–ano*, *–ane* e *–one* apresentaram nessas línguas. Os dados para análise dessas formas no português foram extraídos de dois corpora: 1) Do "Frequency Dictionary of Portuguese Words", um dicionário que contém as 5.000 palavras mais freqüentes no português de Portugal; 2) Do corpus do português culto falado no Brasil, que contém textos de programas de televisão e artigos de revistas brasileiras. Os dados para a análise de outras línguas foram tirados de dicionários das referidas línguas. No caso do português, os dados em *–ão* encontrados pela autora foram os seguintes: no "Frequency Dictionary of Portuguese Words", havia 318 nomes em *–ão*, dos quais 303 eram pluralizados em *–ões* (95,28%), 10 em *–ãos* (3,14%) e cinco em *–ães* (1,57%); no outro corpus (de dados da mídia) havia 227 nomes em *–ão*, sendo 222 pluralizados em *–ões* (97,80%), cinco em *–ãos* (2,20%) e nenhum em *–ães*.

A autora menciona que, desde o português arcaico, a forma *–ões* já era a preferida para a pluralização dos nomes em *–ão* no singular. Com a comparação dos diferentes corpora, a autora observa que os nomes pluralizados em *–ãos* e *–ães* eram basicamente os mesmos nas duas listas comparadas, ao passo que os itens em *–ões* (em maior número) tendiam a variar bastante. A partir desses dados, a autora conclui que as classes de plurais em *–ãos* e *–ães* não são produtivas, visto que, ao longo de vários séculos, não houve (nos corpora consultados) itens novos flexionados com esses morfemas. Por outro lado, o grupo de plural em *–ões* está sempre aumentando, uma vez que, em geral, novas palavras introduzidas na língua são flexionadas com essa terminação.

Os plurais em *–ãos* e *–ães* são infreqüentes, mas, em geral, itens como "cães", "irmãos", "mãos", "pães" tendem a preservar suas formas. Na tentativa de justificar por que tais itens não se conformam ao paradigma mais freqüente (pluralização em *–ões*), a autora levanta as seguintes hipóteses:

- São substantivos primitivos originários do latim e do grego, sendo, portanto, muito antigos e estáveis na língua, em oposição aos nomes pluralizados em *–ões*, que são mais recentes na língua;
- Nomeiam substantivos concretos, de uso familiar e freqüente na vida cotidiana;
- Muitas dessas palavras são monossílabos tônicos, o que as torna relativamente imunes à mudança.

Comparando os dados do português com os das demais línguas românicas, a autora observou que nossa língua foi a única que padronizou as formas *–ano*, *–ane* e *–one* no singular, transformando as três em *–ão*. As demais línguas, cada uma à sua maneira, ainda preservam a distinção etimológica no singular.

A autora realiza, também, uma investigação sobre a morfologia das diferentes terminações de plural em *–ão*. Ela observa que, em geral, *–ãos* e *–ães* fazem parte do radical da palavra, no entanto *–ões* é um sufixo produtivo, apresentando vários significados e origens diferentes. O mais freqüente desses sufixos é *–ção*, que constitui 59,43% de todas as palavras em *–ão* sob análise e é adotado para formar substantivos a partir de verbos, como em "aclamação", "comunicação", "coleção", entre outros. O morfema *–ões* pode representar, também, outros sufixos, mas esses não são tão numerosos na língua. A partir dessa análise, a autora conclui o seguinte:

> Caberia, então, aqui fazer a oposição entre dois tipos de flexão para os nomes em *–ão* do português: uma petrificada, que apenas conserva os tipos antigos, incluindo os nomes cujo plural é em *–ães* e *–ãos* e uma flexão viva, em *–ões* na qual se incluem todos os nomes novos que vão sendo introduzidos na língua, independentemente de sua origem. (CAMPOS, 1981, p. 155)

A análise de Campos (1981) corrobora as informações diacrônicas que apresentamos no capítulo anterior. A autora também afirma que, dentre os morfemas utilizados na pluralização de *–ão*, apenas *–ões* é produtivo. Os demais (*–ãos* e *–ães*) são mais antigos e sua aplicação não se estende a novos itens que entram no PB. Dessa forma, confirma-se a prevalência diacrônica e sincrônica do morfema *–ões* em relação aos demais.

Naro (1981) analisa aspectos da concordância verbal e nominal do PB, investigando as dimensões sociais e estruturais que interferem nesse fenômeno. Quanto à concordância nominal, o autor retoma os dados de Scherre (1978, p. 79), para o efeito da saliência fônica na concordância nominal no Rio de Janeiro, e comenta o seguinte:

> […] there is no relationship between frequency of occurrence of a paradigmatic class and the corresponding rate of agreement. In the educated group, for example, the most frequent class (1a [de plurais regulares, com o acréscimo de *–s*], with 3365 tokens) has the lowest rate of agreement, while the least frequent class (2b [de plurais em metafonia], with only 21 tokens) shows categorical agreement in the sample. (NARO, 1981, p. 96).

O autor afirma que a mesma correlação entre freqüência de classe e taxa de concordância é encontrada para os verbos. Concluindo suas colocações sobre freqüência, o autor observa:

> Although I do not have data for individual items on the lower extreme of the hierarchy [de saliência fônica], the results so far available make it seem unlikely that further investigation would reveal a correlation between frequency of occurrence and rate of agreement on either the paradigmatic or individual item levels. (NARO, 1981, p. 96)

A ênfase no detalhamento das citações de Naro (1981) deve-se ao fato de que esse é um dos únicos trabalhos sobre concordância no PB em que encontramos informações sobre a freqüência da classe paradigmática ou do item lexical propriamente dito. Como vimos, o autor nega que efeitos de freqüência individual ou de classe paradigmática possam ser encontrados na concordância nominal de número do PB. Essa questão sobre freqüência será retomada nas subseções finais deste capítulo.

Scherre e Naro (1992) abordam o tema concordância dentro do sintagma nominal, entre sujeito e verbo e entre sujeito e predicado. Nesse artigo, eles defendem a hipótese de que efeitos de discurso (processamento paralelo) representam o fator mais relevante na presença ou ausência da concordância, porque existe a tendência lingüística de formas similares ocorrerem juntas no discurso, ainda que a marcação nessas formas resulte na redundância dentro da frase.

Avaliando a interação entre processamento paralelo e anos de escolaridade para a concordância entre dois sintagmas nominais, os autores concluíram, basicamente, que: 1) No caso de itens isolados ou de primeiros itens em uma série (em que o processamento paralelo não se aplica, por falta de sintagma precedente), a escolaridade influenciava, visto que a marca de concordância nesses casos aumentou nos indivíduos que freqüentaram a escola por mais anos; 2) No caso da interação entre dois sintagmas nominais, a concordância do segundo sintagma foi menor quando os informantes tinham mais anos de escolaridade, o que demonstra o forte efeito do processamento paralelo, que faz com que o primeiro sintagma nominal influencie o segundo, não permitindo nenhum efeito da escolarização. Os autores justificam os resultados da seguinte maneira:

> Where the linguistic system exerts a high degree of pressure (i.e., in serial tokens under the conditions of parallel processing), the school system has less success; where the linguistic system is less constrained (i.e., in the case of isolated tokens or first occurrences), the school system has more success in increasing the chances of concord use by the speaker. (SCHERRE; NARO, 1992, p. 5)

Scherre (2001a) retoma seus dados de 1988, enfatizando que as variáveis que mais interferem na concordância nominal de número no PB são: 1) Posição linear e relativa dentro do sintagma nominal; 2) Marcas precedentes; 3) Saliência fônica. Segundo a autora, as seguintes generalizações podem ser feitas para esses fatores:

> […] (1) nominal elements to the left of the nominal head favor overt plural markers, while nominal elements to the right of the nominal head favor zero plural markers. The head itself favors overt plural markers if it is more to the left in the noun phrase; otherwise the head disfavors markers; (2) preceding markers lead to more markers, and preceding zeros lead to more zeros; (3) more phonic material differentiation in the singular/plural opposition and stressed final syllable also favor overt plural markers. (SCHERRE, 2001a, p. 2)

Sobre posição dentro do sintagma nominal, Scherre (2001a) afirma que, diferentemente de abordagens anteriores sobre o tema, que concluíram que o primeiro

elemento do sintagma era o mais marcado, a conclusão correta, de acordo com seus dados, é que "all prehead elements show a very high rate of presence of plural markers, while all post-head elements are less marked." (SCHERRE, 2001a, p. 3)

Então, de acordo com as conclusões alcançadas pela autora, em um sintagma nominal como "Todos os meus livros novos importantes", em que o núcleo é "livros", os itens "todos", "os" e "meus" apresentariam maior probabilidade de preservação da concordância, visto que estão antes do núcleo; seguindo a mesma lógica, os itens "novos" e "importantes", por estarem depois do núcleo, apresentariam mais propensão ao apagamento das marcas formais de plural.

Quanto às marcas precedentes, a autora defende o efeito do paralelismo lingüístico, que faz com que marcas levem a marcas e zeros levem a zeros. Nos exemplos "Os meus velhos amigos" e "Umas borrachaø brancaø grandeø", o fato de "os", "meus" e "velhos" apresentarem marcas explícitas de plural aumenta a probabilidade de "amigos" (que vem depois) ser pluralizado também; por outro lado, o fato de "borracha" estar no singular aumenta a probabilidade de "branca" e "grande" (que vêm depois) não serem flexionados. O fator marcas precedentes de plural foi um dos que a autora considerou potencialmente relevantes no processo de perda ou manutenção da concordância nominal de número.

Com relação à saliência fônica, Scherre (2001a) observou que, quando há maior diferenciação de material fônico entre a forma singular e a plural (como [o]vo/[ɔ]vos), as marcas de plural tendem a ser preservadas, ao passo que, quando há pouca diferenciação entre o material fônico do singular e do plural (como "mesa"/"mesas"), a marca de pluralização tende a ser apagada. A autora justifica esse fenômeno da seguinte forma: "[…] forms that are more salient are more perceivable, therefore they receive overt plural marker." (SCHERRE, 2001a, p. 9)

Scherre (2001b) retoma a questão da influência das marcas precedentes na concordância nominal de número. Dessa vez, ela compara dados do PB e do espanhol falado em Porto Rico. Os resultados apontam que a influência do fator marcas precedentes na concordância nominal foi bastante semelhante para as duas línguas analisadas, o que leva a autora a concluir o seguinte:

> I claim that the constraint of preceding markers across linear position on noun phrase number agreement is a parallelism effect at phrase level, undoubtedly another manifestation of linguistic parallelism. This effect is naturally embedded

> in a universal principle of linguistic use: parallel processing (see Scherre & Naro, 1991). Surely parallel processing reflects one of several ways in which the human mind works. My analysis of the parallelism effect on phrase level demonstrates that parallel processing operates more efficiently when preceding markers are more similar. Specifically, if the immediately preceding markers are zeroes, this principle increases in efficiency. (SCHERRE, 2001b, p. 103-104)

Finalizados os comentários sobre os principais trabalhos que a literatura lingüística apresenta sobre concordância nominal e grupos de plurais do PB, passaremos à próxima subseção, que analisa de forma crítica essas pesquisas e descreve uma abordagem alternativa para os plurais em *–ão*, *–l* e ditongo em *–u* do PB.

3.2 Análise crítica da literatura lingüística sobre concordância nominal de número e grupos de plurais do português brasileiro

Quanto às abordagens sobre a concordância nominal de número e grupos de pluralização apresentadas nas subseções anteriores, devemos fazer algumas considerações:

1. As duas análises mais aprofundadas sobre o tema (GUY, 1981; SCHERRE, 1988) adotaram como escopo teórico a Teoria da Variação e Mudança. Essa teoria apresenta uma contribuição importante para a análise lingüística, que é a busca de correlatos entre fenômenos lingüísticos e fatores sociais, tais como gênero, faixa etária, escolaridade e classe social. No entanto, tal teoria não analisa a variação e a mudança lingüística a partir da organização e processamento dos itens no léxico mental dos falantes, já que sua proposta teórica enfatiza mais a correlação entre língua e sociedade;
2. Nas análises anteriores sobre a concordância no PB, observamos que questões como a unidade de armazenagem lexical (palavra inteira ou fonemas e morfemas isolados), possíveis interconexões entre os itens léxicos estocados e como essas informações interferem no processamento lingüístico não foram analisadas, porque esse não era o objetivo principal dos autores. Nessas pesquisas, não se discutiu,

também, a interferência do uso cotidiano da linguagem na forma como as representações mentais de informações lingüísticas estão estruturadas;

3. Nos trabalhos anteriores sobre o tema, observamos que a ênfase era na análise da presença ou da ausência de marcas formais de plural. Com exceção de Campos (1981), não houve nenhuma abordagem que considerasse grupos de plurais específicos, fenômenos recorrentes em tais classes e como a representação mental de informações lingüísticas interfere nesses processos;
4. Não houve, entre os trabalhos resenhados, nenhum que analisasse as variações de plurais que ocorrem nos grupos em *–ão*, *–l* e ditongo em *–u*. Tais classes foram consideradas dentro do contexto mais geral da pluralização do PB, mas não mereceram tratamento específico e individualizado em nenhuma das pesquisas anteriores. Questões como generalizações de plural foram mencionadas em análises precedentes (GUY, 1981; SCHERRE, 1988), no entanto não houve investigações detalhadas a fim de observar os fatores que poderiam interferir na existência de tais variações nesses plurais.

A partir dos comentários acima, consideramos que os grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u* carecem de uma análise sob o foco das representações lingüísticas, com a intenção de investigar como a organização do léxico mental interfere nos processos de pluralização. Essas classes de plurais foram escolhidas como objeto de estudo porque, nesses casos, a pluralização tem outra perspectiva de análise, visto que, além da marcação/não-marcação de plural, há formas concorrentes de plural para a mesma palavra, conforme mencionamos no Capítulo 2, "Grupos de Plurais".

Para o tratamento dessas classes de plurais, esta tese apresenta uma inovação quanto às abordagens anteriores sobre o tema. Nas pesquisas que já foram feitas sobre concordância nominal de número no PB, em geral, casos como "anciãos"/ anciães"/"anciões", e "degraus"/"degrais" eram codificados como marcas de plural para cada uma das classes a que a palavra original pertencia. Apesar da variabilidade aparente, nenhum dos trabalhos anteriores apresentou tratamentos individuais para esses grupos de plurais, tentando justificar a flutuação existente em suas pluralizações. O próprio Guy (1981, p. 181-182) comenta como esse tipo de dado foi codificado:

> For purposes of our analysis all words [terminadas em –ão] were assumed to be classified as they are in the standard dialect, unless they were produced in a leveled form, in which case they were classified as regular. […] The same procedure was followed of classifying words as l–plurals according to their standard dialect comportment, unless they were produced in a levelled (*sic*) form (such as /legaws/ for legais).

Assim, conforme podemos concluir a partir da citação acima, Guy classificou todas as formas variantes de *–l* e *–ão* como pertencentes à língua padrão ou à forma regularizada, sem considerar as peculiaridades de cada um dos plurais adotados nesses grupos. Da mesma forma, não encontramos em Scherre (1988) nenhum tratamento específico para essas formas variantes de plural[29]. A opção de codificação dela é semelhante à de Guy (1981), já que itens com pluralização regular (por exemplo, "irmãos" e "órgãos") são classificados como palavras regulares, e palavras pluralizadas em *–ães* ou *–ões* (como "pães" e "leões") foram alocadas em outra categoria específica. Em Guy (1981) e Scherre (1988), os itens terminados em ditongo em *–u* foram classificados como palavras regulares, pluralizadas em *–s*; portanto, esse grupo também não foi analisado separadamente por esses dois autores.

Dessa forma, percebemos que existe, na literatura lingüística sobre concordância nominal de número no PB, uma lacuna, porque todos os grupos de plurais tiveram, até então, tratamento semelhante, apesar das peculiaridades existentes em cada um deles. Assim, nossa tese propõe uma análise específica para os grupos de plurais em *–ão*, *–l* e ditongo em *–u*, tentando entender as variações existentes nessas classes e suas possíveis justificativas dentro das teorias de representações mentais.

A variabilidade nesses grupos de plurais (de palavras terminadas em *–ão*, em *–l* e ditongo em *–u*) pode fornecer indícios sobre como o uso das palavras no cotidiano do falante afeta as representações mentais de itens léxicos e como a gramática é reconstruída no dia-a-dia de uma comunidade. Além disso, a análise de tais grupos de plurais pode prover mais informações sobre como o léxico mental do falante se organiza em termos de categorização e acesso aos itens lingüísticos. Devemos salientar, também, que as peculiaridades dos grupos de plurais em *–ão*, *–l* e ditongo em *–u*, apesar de não terem sido estudadas separadamente, já foram noticiadas por outros autores, como Guy (1981). Vamos transcrever abaixo alguns de seus comentários sobre esses plurais:

---

[29] Os dados analisados por Guy (1981) e Scherre (1988) não apresentavam quantidade de itens terminados em *–ão*, *–l* e ditongo em *–u* que permitisse a análise desses grupos de plurais em subcategorias.

> We find hypercorrections, such as irmões, and levellings, such as nações, in the data for several informants. Even the standard language has a number of –ão words for which more than one plural ending is possible, such as aldeãos–aldeães–aldeões and verãos–verões. So in our treatment of this category we must bear in mind that its membership, and mental representation, may vary from person to person. (GUY, 1981, p. 181)

Mais adiante, o autor apresenta outro comentário importante sobre o grupo de palavras terminadas em *–l*:

> Another category consisted of words ending in orthographic –l, which in the dialect under consideration [do Rio de Janeiro] is phonetically –w. […] As is the case with –ão plurals, individuals may vary considerably in their treatment of these words. (GUY, 1981, p. 182)

Observe-se, portanto, que Guy (1981) já noticiava a variação nas formas de plurais que esses grupos (em *–ão* e *–l*) apresentam. O autor considera, inclusive, a possibilidade de os itens terminados em *–ão* variarem sua representação mental de indivíduo para indivíduo. No entanto, essa questão não é aprofundada por Guy (1981), porque seu objetivo principal era analisar a presença ou ausência das marcas de plural no PB, não fenômenos específicos registrados em determinadas classes de pluralização. Por conta de todos esses argumentos apresentados anteriormente, defendemos que os grupos de plurais em *–ão*, *–l* e ditongo em *–u* no PB necessitam de uma análise lingüística específica, tendo como foco de análise teorias de representação e estocagem lexical.

Para fornecer maiores detalhes sobre a abordagem alternativa proposta por esta tese, organizamos uma subseção que apresenta informações sobre outras possibilidades de análise que podem ser sugeridas para os grupos de plurais do PB.

3.3 O papel da freqüência na concordância nominal de número do português brasileiro

Conforme veremos no Capítulo 4, "Quadro Teórico", o Modelo de Redes (BYBEE, 1985, 1995, 2001) e a Teoria dos Exemplares (PIERREHUMBERT, 2000, 2001a, 2001b, 2001c) afirmam que a experiência afeta as representações lingüísticas. Em linhas gerais, segundo essas teorias, a unidade básica de estocagem lexical é a palavra inteira

ou até mesmo expressões comumente usadas ("chunks"). Nesse sentido, a freqüência de uso desempenha papel fundamental nas representações mentais, visto que palavras freqüentes são mais disponíveis no léxico mental e podem, inclusive, adquirir certa independência funcional; por outro lado, palavras pouco freqüentes são recuperadas no léxico mental com mais dificuldade e, por isso, são mais suscetíveis a fenômenos como nivelamento analógico. Segundo Bybee (2002a), existem dois tipos de medidas de freqüência: ocorrência ("token") e tipo ("type"). Tais modalidades de freqüência são explicadas pela autora da seguinte maneira: "Token or text frequency is the frequency of occurrence of a unit, usually word or morpheme, in running text. [...] Type frequency refers to the dictionary frequency of a particular pattern, such as stress pattern, an affix, etc." (BYBEE, 2002a, p. 604). Um exemplo de freqüência de ocorrência seria contabilizar quantas vezes a palavra "avião" ocorre em um corpus do PB, ao passo que um exemplo de freqüência de tipo seria contar quantas vezes o sufixo *–vel* ("amável", "agradável", "compatível") ocorre no dicionário ou em um corpus do PB. Nesse sentido, como as representações mentais são baseadas (entre outros fatores) em freqüência, tanto de tipo quanto de ocorrência, é importante que, ao analisarmos um determinado fenômeno lingüístico, seja ele fonológico, morfológico, sintático ou semântico, essas informações de freqüência sejam consideradas. Não mencionaremos, neste ponto, maiores detalhes relacionados às noções de freqüência de tipo e de ocorrência e suas implicações para o léxico mental, porque esse assunto será minuciosamente abordado no Capítulo 4, "Quadro Teórico". Neste momento, nosso objetivo é apenas tecer algumas considerações iniciais sobre o papel da freqüência nas representações lingüísticas. Por conta disso, nos próximos parágrafos, faremos alguns breves comentários sobre a freqüência de tipo e de ocorrência dos grupos de plurais sob análise nesta tese.

Com relação ao plural das palavras terminadas em *–ão* no singular, a grande fonte de divergência nesse grupo é o fato de que existem, potencialmente, três possibilidades de plural (*–ões*, *–ãos* e *–ães*) para um único singular (*–ão*). Por conta disso, algumas vezes um plural é utilizado quando outro seria o adequado, etimologicamente falando.

Em meio a toda essa aparente controvérsia, as informações diacrônicas apresentadas no Capítulo 2, "Grupos de Plurais", demonstram que as variações nos itens terminados em *–ão* não são aleatórias, mas apresentam uma direção específica: as palavras

que apresentam plurais etimológicos[30] em *–ãos* ou *–ães* generalizam-se, por analogia, para o grupo de plurais em *–ões*, não o contrário (palavras com plural etimológico em *–ões* raramente se generalizam para *–ãos* ou *–ães*[31]). Observamos, no capítulo anterior, que informações diacrônicas sobre a evolução das formas latinas para o PB comprovam tal hipótese. A primeira explicação para essa prevalência do morfema *–ões* pode ser o fato de que, aparentemente, a terminação plural *–ões* é mais freqüente que as outras duas (*–ãos* e *–ães*). A fim de obter informações precisas para confirmar ou retificar essa hipótese, fizemos uma busca no Corpus ASPA[32] e observamos as seguintes freqüências de morfemas (tipos) de plural para as palavras em *–ão* no singular:

Tabela 5: Freqüência dos tipos terminados em *–ão* (plural) no ASPA

Tipos	**N**	**% no grupo *–ão***
–ões	1.352	97,3
–ãos	25	1,7
–ães	12	0,8
Total	**1.389**	**100**

A partir da tabela acima, percebemos que, realmente, o plural em *–ões* é mais freqüente (97,3%) que os tipos em *–ãos* e *–ães* (1,7% e 0,8 %, respectivamente) e, a princípio, essa pode ser uma fonte de explicação para a direção que as generalizações de plural seguem (de *–ãos* e *–ães*, plurais infreqüentes, para *–ões*, plural mais freqüente). Conforme veremos no Capítulo 4, "Quadro Teórico", as representações mentais dos itens léxicos são reorganizadas a partir do uso da língua em situações concretas de comunicação. Nesse sentido, itens pertencentes a um paradigma menos freqüente podem migrar para um padrão mais freqüente, o que parece ser o caso dos plurais em *–ão* do PB.

O ASPA lista 25 palavras pluralizadas em *–ãos* e 12 em *–ães*. No entanto, pode-se questionar que, como esse corpus é baseado em dados de uso da língua escrita e não é um dicionário, não se pode afirmar que no PB existem apenas 37 itens pluralizados em *–ãos* e *–ães*. Essa crítica é pertinente, no entanto devemos lembrar que o Modelo de Redes e o

[30] Conforme já esclarecemos anteriormente, nesta tese, adotaremos os termos "plural etimológico" para os plurais originais que eram usados no latim e "plural analógico" para casos de palavras que desenvolveram, no PB, outro plural, diferente do etimológico.
[31] Conforme mencionamos anteriormente, os itens "bênção", "peão" e "pião" foram os únicos exemplos encontrados de itens de etimologia *–ões* que atualmente adotam plural em *–ãos* ou *–ães*.
[32] No Capítulo 5, "Metodologia", apresentaremos maiores informações sobre esse corpus.

de Exemplares, suportes teóricos desta tese, conferem papel importante ao uso das formas lingüísticas na organização do léxico mental dos falantes. A contagem das palavras no dicionário da língua é importante, mas não garante que todos os itens listados sejam, efetivamente, utilizados no dia-a-dia de uma comunidade de fala. Dicionários contêm formas arcaicas e obsoletas que são, por vezes, totalmente desconhecidas pelos falantes, por isso um modelo de léxico mental que tem como pilar o pressuposto de que a experiência (uso concreto das formas lingüísticas) afeta as representações mentais não pode pautar sua contagem de freqüência considerando apenas dados extraídos de dicionários. A utilização de um corpus de língua (em vez de um dicionário) pode fornecer índices mais precisos sobre o uso real das formas lingüísticas no cotidiano do falante. Apesar dessas considerações, apenas para efeito de comparação com os dados do ASPA, consultamos o Novo Dicionário Aurélio Eletrônico para verificar a quantidade de itens que se flexionava através dos padrões em *–ões*, *–ãos* e *–ães*. Apresentamos abaixo os resultados dessa busca:

Tabela 6: Freqüência dos tipos terminados em *–ão* no Novo Dicionário Aurélio Eletrônico

Tipos	**N**	**% no grupo *–ão***
–ões	5.720	97,5
–ãos	95	1,6
–ães	47	0,8
Total	**5.862**	**100**

Comparando as Tabelas 5 e 6 acima, observamos que os dados de língua escrita do ASPA e os verbetes do Novo Dicionário Aurélio Eletrônico apresentam quantidades diferenciadas de itens terminados em *–ões*, *–ãos* e *–ães*. Em geral, um dicionário registra, além de itens freqüentes, formas pouco usadas na língua cotidiana, por isso é natural que apresente um maior quantitativo de palavras. Portanto, essa diferença numérica é normal, em se tratando da comparação entre um corpus de língua escrita e um dicionário. No entanto, analisando as porcentagens de cada morfema dos plurais em *–ão* (em relação ao grupo em *–ão* como um todo), observamos que houve percentuais praticamente idênticos nos dois corpora consultados: no ASPA, *–ões* tem 97,3% de dados e, no Aurélio, 97,5%; no ASPA, *–ãos* tem 1,7% de palavras e, no Aurélio, 1,6%; no ASPA e no Aurélio, *–ães* tem 0,8% de itens. Concluímos, portanto, que a prevalência do tipo em *–ões* se confirma nos dois corpora analisados. Isso ratifica nossa hipótese de que os dados

extraídos de corpora de uso da língua são bons indicadores para a freqüência de tipo dos itens léxicos, apesar de não listarem todas as palavras do dicionário. No caso dos plurais em *–ão*, os morfemas *–ãos* e *–ães* apresentam baixa freqüência de ocorrência e talvez por isso estejam migrando para *–ões*, conforme vimos na análise diacrônica apresentada no Capítulo 2, "Grupos de Plurais".

Com relação à freqüência de tipo dos plurais *–ãos*, *–ães* e *–ões* no português arcaico, Mattos e Silva (19--) apresenta a seguinte estatística de itens em *–ão* (no singular ou no plural) no corpus do português do século XIV:

Tabela 7: Quantidade de itens terminados em *–ão* no corpus do português do século XIV

Tipos	**Exemplos**	**Quantidade de itens**	**% no total do corpus**
–ã + VT[33] *o* (sing. e pl.[34])	cidadão ~ cidadãos	121	18
–ã + VT *e* (sing. e pl.)	pan ~ pães	75	11,1
–õ + VT *e* (sing. e pl.)	baron ~ barões	476	70,8

Na tabela acima, observamos que, no português do século XIV, o morfema *–ões* já era o mais freqüente, seguido por *–ãos* e *–ães*, nessa ordem. A partir dos dados coletados para o português atual (Cf. Tabelas 5 e 6 acima), percebemos que essa tendência se mantém até hoje. Há, portanto, evidências diacrônicas e sincrônicas que apontam para a preferência dos falantes pelo morfema *–ões* para pluralizar itens terminados em *–ão* no singular. É importante enfatizar, no entanto, que essa alta freqüência de tipo do morfema *–ões* deve ser analisada em função da sua interface com freqüência de ocorrência, da seguinte maneira: formas em *–ãos* e *–ães* são suscetíveis a adotar plurais em *–ões*, por causa da alta freqüência de tipo desta última; no entanto, no grupo de *–ãos* e *–ães* há itens de alta freqüência de ocorrência, como "alemães" (1.025 ocorrências no Corpus NILC/São Carlos[35]) e "irmãos" (1.125 ocorrências no mesmo corpus), cujas formas etimológicas talvez sejam preservadas por conta de sua recorrência no uso cotidiano. Essa hipótese será verificada em nossos dados.

Nos parágrafos seguintes, comentaremos informações sobre freqüência de tipo e de ocorrência para os plurais terminados em *–l* e ditongo em *–u*.

---

[33] VT deve ser entendido como "vogal temática".
[34] A autora não contabiliza em separado as formas de singular e plural.
[35] Maiores detalhes sobre esse corpus serão mencionados na subseção 5.3 do Capítulo "Metodologia".

No Capítulo 2, "Grupos de Plurais", mencionamos o fato de que *–l* em fim de sílaba, na maioria dos dialetos do PB, é geralmente pronunciado como [w]. A distinção fonética entre *–l* e *–u* em fim de sílaba no PB se perdeu e generalizações podem ser encontradas nesses grupos, no sentido de a pluralização em *–l* ser aplicada quando a palavra termina em *–u* (ortográfico) ou vice-versa. Nossa hipótese, a partir de intuições lingüísticas, é de que, em geral, as palavras terminadas em *–u* estão migrando para a classe em *–l*, mas o contrário não é muito comum. A fim de encontrar justificativas para essa hipótese, fizemos uma busca pelas palavras terminadas em *–l* e ditongo em *–u*. No ASPA, encontramos as seguintes freqüências de tipo para tais grupos:

Tabela 8: Freqüência dos tipos terminados em *–l* (singular e plural) no ASPA

Tipo	**Freqüência de Singular**		**Tipo**	**Freqüência de Plural**	
	N	**% no grupo *–l***		**N**	**% no grupo *–l***
–al	972	44,1	*–ais*	563	82,9
–el	847	38,4	*–eis*[36] / *–éis*	34	5
–il	220	9,9	*–is*	64	9,4
–ol	121	5,4	*–óis*	15	2,2
–ul	41	1,8	*–uis*	3	0,44
Total	**2.201**	**100**	**Total**	**679**	**100**

Tabela 9: Freqüência dos tipos terminados em ditongo em *–u* (singular e plural) no ASPA

Tipo	**Freqüência de Singular**		**Tipo**	**Freqüência de Plural**	
	N	**% no grupo *–u***		**N**	**% no grupo *–u***
–au	81	45	*–aus*	21	72,4
–eu / *–éu*	72	40	*–eus* / *–éus*	08	27,5
–iu	15	8,3	*–ius*	0	0
–ou	12	6,6	*–ous*	0	0
–uu	0	0	*–uus*	0	0
Total	**180**	**100**	**Total**	**29**	**100**

A partir da análise das tabelas acima, podemos perceber que o PB apresenta mais tipos em *–l* do que em ditongo em *–u*. Com relação às terminações em *–iu*, *–ou* e *–uu*,

[36] Todos os plurais de *–el* registrados são em –[ɛ]*is*. Não foi encontrado nenhum caso de –[e]is.

a diferença é ainda maior, porque o número de itens que apresentam tais terminações é baixo (27 no singular e zero no plural[37]).

O mesmo questionamento sobre a freqüência no ASPA dos itens terminados em *–ão* é cabível quanto àqueles em *–l* e ditongo em *–u*. O ASPA é um corpus que reflete o uso da língua, mas não apresenta a ocorrência de todas as palavras do PB. Esse tipo de informação só pode ser obtido através da consulta a um dicionário. A fim de investigar essa possível crítica, fizemos, no Novo Dicionário Aurélio Eletrônico, uma busca pelas palavras terminadas em *–l* e ditongo em *–u* e organizamos tabelas com essas estatísticas, da mesma forma como procedemos para os plurais em *–ão*. Como o dicionário lista os verbetes sem suas flexões plurais, apresentamos apenas a freqüência singular desses grupos. Os resultados podem ser visualizados nas tabelas abaixo:

Tabela 10: Freqüência dos tipos terminados em *–l* no Novo Dicionário Aurélio Eletrônico

Tipos	**N**	**% no grupo *–l***
–al	2.506	50,1
–el	1.846	36,9
–il	297	5,9
–ol	267	5,3
–ul	83	1,6
Total	**4.999**	**100**

Tabela 11: Freqüência dos tipos terminados em ditongo em *–u* no Novo Dicionário Aurélio Eletrônico

Tipos	**N**	**% no grupo *–u***
–au	147	35,3
–eu / *–éu*	241	57,9
–iu	09	2,1
–ou	12	2,8
–uu	07	1,6
Total	**416**	**100**

[37] O Novo Dicionário Aurélio Eletrônico registra os seguintes substantivos e adjetivos terminados em *–iu*, *–ou* e *–uu*: "abiu", "ararapitiu", "djidiu", "pipiu", "psiu", "quem-te-vestiu", "tiu", "tiziu", "xibiu"; "apeou", "atouaou", "clou", "flou", "fogo-apagou", "fogo-pagou", "gente-de-fora-já-chegou", "grou", "menejou", "ouranaiou", "samou", "touaou"; "anuu", "inambuu", "inamuu", "inhambuu", "nambuu", "nhambuu", "Uuu". No entanto, no ASPA, fonte dos dados das tabelas acima, essas palavras não foram registradas.

Comparando os dados do ASPA e os do Novo Dicionário Aurélio Eletrônico para os dois grupos de plurais, observamos que existem diferenças numéricas no quantitativo de palavras para cada um dos corpora; no entanto, percentualmente a distribuição das ocorrências é bastante similar, tal como aconteceu no grupo terminado em *–ão*. Considerando as formas de singular do ASPA e do Novo Dicionário Aurélio Eletrônico[38], podemos observar que:

- Na classe de palavras em *–l* (Cf. Tabelas 8 e 10 acima), no ASPA e no Aurélio existe exatamente a mesma gradação de freqüência de tipo para cada um dos morfemas: *–al* é o mais freqüente, *–el*, *–il*, e *–ol* (nesta ordem) vêm a seguir e *–ul* é o menos freqüente;
- Na classe de palavras em ditongo em *–u* (Cf. Tabelas 9 e 11 acima), *–au* é o mais freqüente no ASPA e o segundo mais freqüente no Aurélio; *–eu* é o mais freqüente no Aurélio e o segundo mais freqüente no ASPA; nos dois corpora, *–iu*, *–ou* e *–uu* apresentam baixa quantidade de dados.

Observamos, portanto, que os dados percentuais advindos dos dois corpora (ASPA e Novo Dicionário Aurélio Eletrônico) confirmam a informação de que, tanto na língua escrita corrente quanto no dicionário, o tipo em *–l* é mais freqüente que o tipo em ditongo em *–u* (ambos em posição final de sílaba).

A partir dessas estatísticas, podemos formular uma primeira hipótese para o fato de palavras terminadas em *–u* adotarem a regra de plural para palavras em *–l*, como nos casos de "degrais", em vez de "degraus", e "troféis", em vez de "troféus": por causa da perda de distinção fonética entre *–l* e *–u* em fim de sílaba no PB e da baixa freqüência de tipo do morfema ditongo em *–u*, os falantes, quando precisam pluralizar uma palavra terminada em [w], parecem adotar a regra de pluralização em *–l*, já que esse tipo, tanto no singular quanto no plural, é mais freqüente que o tipo *–u* no fim de palavra. Também nesse caso, uma interação entre as freqüências de tipo e de ocorrência é importante, visto que é possível que algumas palavras em ditongo em *–u* estejam migrando para a classe de plurais em *–l*, mas, no grupo de itens terminados em *–au* e *–eu*, há alguns itens de alta freqüência

[38] Por uma questão de equivalência e compatibilidade dos dados, a comparação foi feita apenas para as formas de singular, já que o ASPA fornece itens em sua forma singular e plural, mas o Novo Dicionário Aurélio Eletrônico apresenta os verbetes sem flexão, ou seja, apenas no singular.

de ocorrência plural, como "grau" (2.770 ocorrências no Corpus NILC/São Carlos) e "museu" (1.054 ocorrências no mesmo corpus). Talvez a alta freqüência de ocorrência desses itens previna-os contra variações analógicas, como o caso da migração de itens em ditongo em *–u* para a classe em *–l*.

Essas questões relacionadas ao uso da língua no cotidiano do falante e ao papel da freqüência como desencadeador de variações lingüísticas definem, de forma mais objetiva, a análise alternativa proposta por esta tese para os plurais terminados em *–ão*, *–l* e ditongo em *–u* do PB. Finalizadas as considerações sobre o papel da freqüência nos processos morfofonológicos de formação de plural no PB, passaremos para a próxima subseção, que encerra este capítulo.

3.4 Conclusões

Comparando as diferentes abordagens sobre concordância nominal e grupos de plurais do PB comentadas nos Capítulos 2 e 3 ("Grupos de Plurais" e "Revisão de Literatura"), observamos que, apesar de a gramática normativa apresentar regras para a concordância nominal de número, variações tendem a ocorrer, sobretudo na língua falada. Informações diacrônicas (evolução das formas de plural do latim até o português) e fatos sincrônicos (freqüência de tipo e de ocorrência dos itens lexicais e morfemas de plural, segundo corpora do PB) demonstram que as controvérsias noticiadas nos grupos de plurais em *–ão*, *–l* e ditongo em *–u* ocorrem porque o uso pode interferir na forma como as palavras são armazenadas e acessadas no léxico mental. Conforme apontamos na subseção anterior, as variações observadas nessas classes de plural talvez sejam justificadas por causa da interferência das freqüências de tipo e de ocorrência nas representações mentais. No Capítulo 6, "Análise de Dados", vamos verificar se a freqüência foi, de fato, relevante para as migrações de plurais nos grupos terminados em *–ão*, *–l* e ditongo em *–u*.

Com as informações apresentadas neste capítulo, concluímos, também, que abordagens anteriores sobre a concordância nominal não analisaram em profundidade questões relacionadas à armazenagem e ao processamento dos itens lexicais, pois utilizaram como escopo teórico a Teoria da Variação e Mudança Lingüística, que busca correlatos

entre variações lingüísticas e fatores sociais, mas não enfoca questões como a representação mental de informações lingüísticas.

Através das estatísticas sobre freqüência de tipo e ocorrência apresentadas na subseção anterior deste capítulo, pudemos vislumbrar diferentes possibilidades para a análise dos grupos de plurais em *–ão*, *–l* e ditongo em *–u* do PB. Freqüência é um dos elementos que interferem na forma como a mente humana categoriza informações (tanto lingüísticas quanto não-lingüísticas), por isso esse fator é importante e precisa ser considerado em uma análise que se propõe a observar relações entre processos morfofonológicos de plural e representações mentais. Um fenômeno de variação e mudança lingüística deve, portanto, avaliar se efeitos de freqüência são relevantes para que ocorram variações nas estruturas lingüísticas.

Finalizadas as considerações acerca dos trabalhos sobre a pluralização nominal do PB como um todo e sobre os grupos de plurais analisados nesta tese, passamos para o próximo capítulo, que expõe, diacronicamente, os principais modelos de léxico mental já propostos na literatura lingüística. Nesse mesmo capítulo, apontamos e justificamos nossa opção teórica pelo Modelo de Redes e pela Teoria dos Exemplares para a análise dos dados coletados sobre os grupos de plurais em *–ão*, *–l* e ditongo em *–u* do PB.

CAPÍTULO 4

Quadro Teórico

Ao longo das últimas décadas, um dos assuntos que mais mereceu a atenção de lingüistas é a questão da estocagem das palavras no léxico mental dos falantes. Diferentes teorias foram propostas no sentido de entender como a mente humana categoriza, armazena e acessa informações lingüísticas (Cf. RUMELHART; McCLELLAND, 1986; MacWHINNEY; LEINBACH, 1991; DAUGHERTY; SEIDENBERG, 1994; PINKER; PRINCE, 1988, 1994; PINKER, 1991, 1999; BYBEE, 1985, 1995, 2001; PIERREHUMBERT, 2000, 2001a, 2001b, 2001c).

Apesar de existirem vários modelos teóricos que discutem esse tema, em linhas gerais, a diferença mais proeminente entre tais abordagens reside no papel que cada uma delas confere ao léxico. De um lado, encontramos propostas, como a Teoria de Palavras e Regras[39], que defendem a existência de regras abstratas a partir das quais palavras e sentenças são formadas. Nesse tipo de modelo, o papel do léxico é relativamente reduzido, porque se propõe a armazenagem de radicais e afixos e, a partir da aplicação de regras abstratas, palavras e sentenças são formadas. Palavras inteiras são armazenadas apenas se forem irregulares e, portanto, não dedutíveis a partir de regras. Por outro lado, existem propostas, como o Modelo de Redes e a Teoria dos Exemplares[40], que conferem papel crucial ao léxico, visto que todas as palavras, inclusive as regulares e morfologicamente complexas, podem ser estocadas e, além disso, existe, também, a possibilidade de armazenagem de fatores pragmáticos e sociais associados aos itens léxicos. Portanto, sob o ponto de vista macroestrutural, poderíamos afirmar que as teorias de estocagem de informações lingüísticas diferem, primordialmente, em sua forma de conceber o léxico mental.

---

[39] Cf. PINKER; PRINCE, 1988, 1994; PINKER, 1991, 1999; PINKER; ULLMAN, 2002a, 2002b.
[40] Cf. BYBEE, 1985, 1995, 2001; PIERREHUMBERT, 2000, 2001a, 2001b, 2001c.

Dentro desse contexto, esta tese faz opção clara por teorias que conferem ao léxico papel imprescindível nas representações lingüísticas, como o Modelo de Redes e a Teoria dos Exemplares. Os principais motivos que sedimentam essa escolha são:

1) A experiência afeta a forma como o conhecimento humano é categorizado (BYBEE, 2001), por isso a freqüência de uso da língua no cotidiano interfere na representação mental de informações lingüísticas;
2) A habilidade lingüística é parte da cognição humana e não apresenta um módulo diferenciado na mente, portanto informações lingüísticas são processadas da mesma forma que outros tipos de conhecimento (BYBEE, 1995, 2001; HOPPER, 1998);
3) O cérebro humano é capaz de armazenar múltiplas informações sobre objetos cognoscíveis, por isso palavras regulares e irregulares podem ser estocadas no léxico mental (PIERREHUMBERT, 2000, 2001a, 2001b; 2001c; JOHNSON, 1997).

Partindo-se desses pressupostos, o objetivo deste capítulo é apresentar uma revisão sobre alguns modelos de léxico já propostos na literatura lingüística. Como nossa tese tem um objeto de estudo específico (efeitos de freqüência nas representações mentais dos plurais em *–ão*, *–l* e ditongo em *–u* no PB), faremos uma resenha sobre as teorias de estocagem lexical mencionadas acima, tentando entender como cada uma delas contribuiria para a análise das variações nos grupos de plurais que são objeto de estudo nesta tese (Cf. Capítulos 2 e 3, "Grupos de Plurais" e "Revisão de Literatura"). Apesar de fazermos opção teórica pelo Modelo de Redes e de Exemplares, discutiremos, neste capítulo teórico, os pressupostos de outros dois modelos de estocagem lexical: o Modelo Conexionista e a Teoria de Palavras e Regras. Nosso objetivo, com essa análise, é fazer uma cronologia das diferentes teorias já propostas sobre a categorização e estocagem lexical, comparando as controvérsias e diferenças existentes entre elas e justificando nossa opção teórica pelo Modelo de Redes e pela Teoria dos Exemplares.

Ademais, este capítulo está organizado da seguinte forma: as subseções 4.1, 4.2, 4.3 e 4.4 apresentam, nesta ordem, informações sobre o Conexionismo, a Teoria de Palavras e Regras, o Modelo de Redes e a Teoria de Exemplares; finalmente, a subseção 4.5 faz uma breve conclusão sobre os modelos de léxico mental discutidos ao longo deste capítulo.

4.1 Conexionismo

Conexionismo ("parallel distributed processing", "neural networks" ou "neurocomputing") é, originalmente, o nome dado a modelos de redes neurais que tentam emular o funcionamento de processos cognitivos. Tais sistemas, em vez de apenas receberem e processarem informações, são capazes de aprender a partir da experiência com os estímulos a que são expostos, estando aptos, portanto, a simular a forma como o cérebro humano armazena e acessa informações lingüísticas:

> Os modelos conexionistas surgiram com o objetivo de simular [a] topologia neurológica do cérebro. Vários elementos computacionais simples, semelhantes a neurônios, são interconectados em redes, recebendo, em paralelo, simultaneamente, sinais uns dos outros através de suas conexões, exercendo uma influência mútua entre si. (CIELO, 1998, p. 45)

Outra característica importante das redes conexionistas é o fato de elas poderem ser arquitetadas com um algoritmo interno que lhes possibilita a aprendizagem de padrões para os quais foi construída:

> Os modelos conexionistas podem ser treinados para desempenhar diversas tarefas, tais como prever o reaparecimento de um objeto por detrás de uma tela, flexionar um verbo no passado, prever a palavra seguinte numa frase, categorizar objetos, categorizar sons de fala, sonorizar um texto escrito, pegar uma bola ou equilibrar uma régua. Em cada caso, o algoritmo de aprendizagem vai ajustando cuidadosamente a força das conexões na rede até alcançar um desempenho similar ao de um adulto. (PLUNKETT, 2000, p. 114-115)

Uma das grandes vantagens dos modelos conexionistas é o fato de ajudarem no estudo da cognição, demonstrando, de forma explícita, como a aprendizagem ocorre dentro das redes. Existem vários modelos de redes neurais, mas, em linhas gerais, pode-se dizer que eles partilham algumas características, a saber (Cf. ELMAN, 1998a):

- Em modelos conexionistas, o processamento é feito por grupos de elementos chamados nós ou unidades. Esses itens são elaborados para imitar neurônios, e o acesso a um desses nós se reflete nos demais;
- Em vez de utilizarem representações simbólicas, as redes empregam padrões de ativação ao longo de diferentes unidades;

- Através de algoritmos de aprendizagem, as redes são capazes, por si mesmas, de atribuir diferentes pesos para as conexões, dispensando, portanto, que um ser humano faça tais ajustes;
- Relacionando formas de entrada (estímulo apresentado) com formas de saída (resultado alcançado), a rede ajusta os pesos das conexões e melhora seu desempenho através de sucessivos treinamentos.

Nos parágrafos seguintes, discutiremos os principais modelos conexionistas já propostos na literatura lingüística. Procuraremos explicar sua arquitetura e apresentar os resultados que eles conseguiram alcançar.

Rumelhart e McClelland (1986) são considerados os pioneiros no desafio de avaliar se uma rede neural poderia aprender padrões similares aos adquiridos pelos seres humanos. Os autores aplicam o Modelo Conexionista ao processo de aprendizagem do passado dos verbos do inglês. Essa flexão foi escolhida porque a regularização que crianças fazem na aquisição dos verbos (aplicando o sufixo *–ed* a verbos irregulares, como "bring"/"bringed" ou "think"/"thinked") é comumente citada como uma evidência da atuação de regras no léxico mental. Como os autores têm a intenção de negar esse postulado teórico (necessidade de regras), escolheram essa flexão para verificar se o modelo conexionista seria capaz de aprender a flexionar verbos do inglês sem usar regras abstratas. A fim de testar a viabilidade dessa hipótese, os autores elaboram uma rede neural com a seguinte estrutura (Cf. RUMELHART; McCLELLAND, 1986, p. 222):

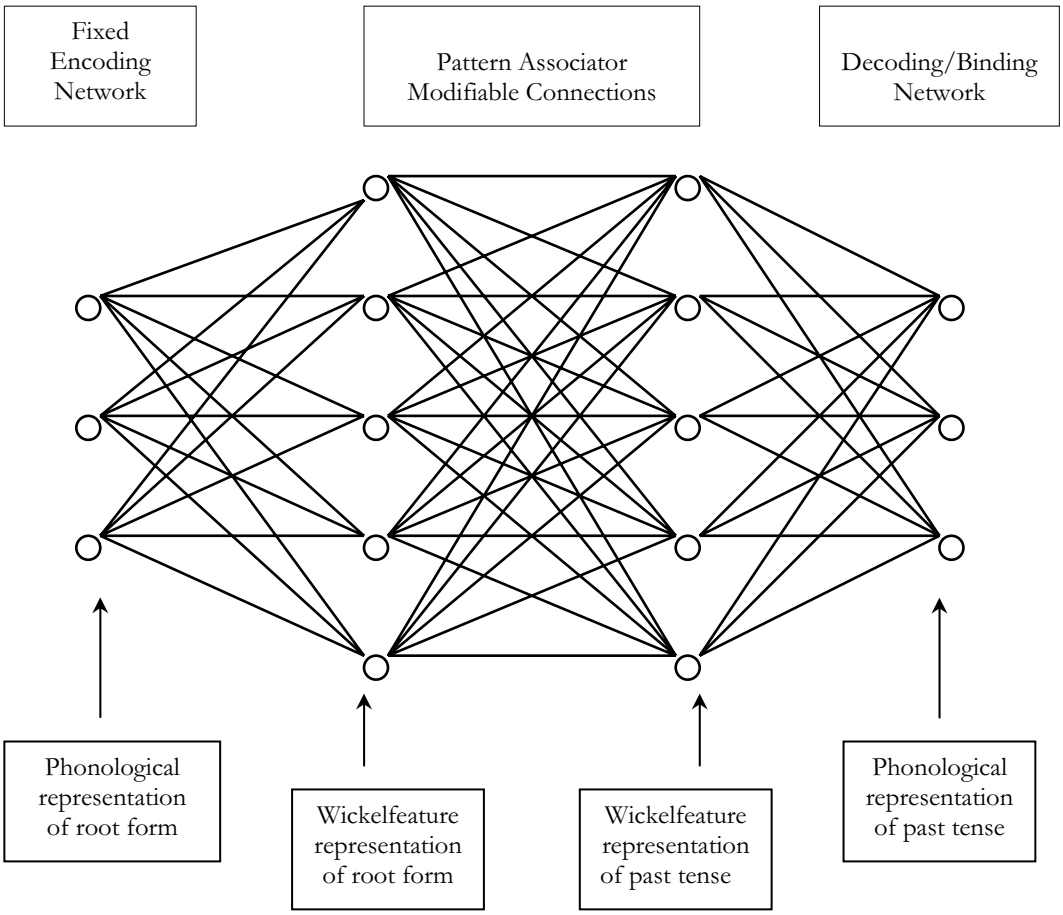


Figura 1: A estrutura básica do Modelo Conexionista de Rumelhart e McClelland (1986)

O modelo proposto por Rumelhart e McClelland (1986) consiste de duas partes básicas:

> (a) a simple pattern associator network [...] which learns the relationships between the base form and the past-tense form, and (b) a decoding network that converts a featural representation of the past-tense form into a phonological representation. All learning occurs in the pattern associator; the decoding network is simply a mechanism for converting a featural representation which may be a near miss to any phonological pattern into a legitimate phonological representation. (RUMELHART; McCLELLAND, 1986, p. 222-223)

No modelo de Rumelhart e McClelland (1986), a representação fonológica dos itens léxicos é feita através de "Wickelphones" e "Wickelfeatures". Essa abstração decorre de um esquema proposto por Wickelgren (1969), segundo o qual cada palavra é representada por unidades designando um fonema, seu antecessor e seu sucessor. Esses seriam os Wickelphones. De acordo com essa estrutura fonológica, quando o radical de um

verbo do inglês é apresentado ao modelo, ele o codifica em Wickelphones que gerarão, através de conexões múltiplas, uma representação de passado, também utilizando essa mesma codificação. Um dos problemas dessa representação é que ela precisa de um número enorme de combinações para conseguir cobrir todas as seqüências possíveis da língua; por causa disso, Rumelhart e McClelland (1986) adotam a noção de Wickelfeatures, representando os traços de cada um dos fonemas, permitindo que uma quantidade menor de informações seja arquivada.

Os modelos conexionistas tentam simular as peculiaridades da mente humana, portanto "[...] they consist of large numbers of simple processors, known as *units* (or nodes), which are densely interconnected into a complex network, and which operate simultaneously and co-operatively to solve information processing problems." (CHRISTIANSEN; CHATER, 2000, p. 4). Dentro dessa perspectiva, o funcionamento básico do modelo se dá da seguinte forma:

> When each input is presented, it is fed through the network, and the output is derived. The output is compared against the correct "target" value and the difference between the two is calculated for each output unit. The squared differences are summed over all the output units, to give an overall measure of the "error" that the network has made. The goal of learning is to reduce the overall level of error, averaged across input/target pairs. (CHRISTIANSEN; CHATER, 2000, p. 4)

Portanto, quando o modelo se depara com uma "entrada" (input), ele produz um "resultado" (output) e este é comparado com um alvo, que é a forma real (gramaticalmente correta) daquela palavra. A diferença entre a forma produzida pelo modelo e a forma esperada é calculada para cada resultado alcançado e, através de ajustes na rede, o sistema vai aprendendo a fornecer as formas corretas para as palavras apresentadas.

O Modelo Conexionista pressupõe a aprendizagem sem a noção de regras simbólicas. Rumelhart e McClelland (1986) esclarecem que o desempenho lingüístico é caracterizado por regras, mas essas não precisam, necessariamente, estar explícitas no mecanismo de aquisição e acesso a informações lingüísticas, ou seja, as regras são uma generalização a partir da realidade lingüística, mas a aprendizagem não se dá, obrigatoriamente, através delas. Como não há regras formais a serem aprendidas, o processamento da linguagem ocorre através do ajuste gradual de conexões feitas entre os nós na recepção de entradas e produção de resultados.

No treinamento do modelo conexionista proposto por Rumelhart e McClelland (1986), foram utilizados 506 verbos extraídos do Corpus Kučera e Francis (1967). As freqüências de ocorrência desses verbos foram consideradas tendo como base a freqüência do gerúndio. Havia 10 verbos de alta freqüência, 410 de média e 86 de baixa freqüência. No primeiro estágio de treinamento, o modelo recebeu os 10 verbos mais freqüentes, sendo oito regulares e dois irregulares. Posteriormente, no decorrer do treinamento, o modelo recebeu os 410 verbos de média freqüência (334 regulares e 76 irregulares) e, mais adiante, os 86 de baixa freqüência (72 regulares e 14 irregulares).

Depois de algum tempo de treinamento, o modelo de Rumelhart e McClelland (1986) obteve 90% de acerto na flexão verbal de passado e foi capaz de produzir resultados não somente para aqueles itens com que foi treinado, mas para qualquer palavra apresentada, pois se tornou apto a generalizar através do mapeamento entre as entradas e os resultados (palavras apresentadas ao modelo e formas de passado produzidas por ele).

Um dos grandes sucessos alcançados pelo Modelo Conexionista na aprendizagem do passado dos verbos do inglês é o fato de ele conseguir reproduzir as três fases do desenvolvimento da criança na aquisição dessa flexão[41]: 1) Utilização de poucos verbos na forma de passado, geralmente os irregulares e de alta freqüência de ocorrência; 2) Regularização de vários verbos segundo o paradigma da desinência *–ed*, inclusive os já aprendidos no estágio anterior; 3) Coexistência dos verbos regulares e irregulares em suas formas corretas, embora regularizações ainda possam ocorrer, inclusive na fase adulta, devido à baixa freqüência de alguns verbos. Conforme mencionamos anteriormente, o Modelo Conexionista seguiu essas mesmas etapas na aprendizagem da flexão de passado dos verbos do inglês. A avaliação dos autores sobre o modelo é a seguinte:

> We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a "rule" as anything more than a description of the language. We have shown that, for this case, there is no induction problem. The child need not figure out what the rules are, nor even that there are rules. The child need not decide whether a verb is regular or irregular. There is no question (as far as generating the past-tense form is concerned) as to whether a verb form is one encountered many times or one that is being generated for the first time. A uniform procedure is applied for producing the past-tense in every case. The base form is supplied as input to the past-tense network and the resulting pattern of activation is interpreted as a phonological representation of the past form of that verb. This is the procedure whether the verb is regular or irregular, familiar or novel. (RUMELHART; McCLELLAND, 1986, p. 267)

[41] O processo de desenvolvimento pelo qual a criança passa na aquisição da flexão verbal do inglês é conhecido como "aprendizagem em forma de 'U'" ("U-shaped learning").

Apesar de bem-sucedido na aprendizagem da flexão de passado do inglês, o modelo de Rumelhart e McClelland (1986) sofreu algumas críticas, dentre as quais podemos citar (Cf. PINKER; PRINCE, 1988, 1994):

1) O fato de o modelo ter sido treinado com um corpus composto primordialmente por verbos irregulares pode ter sido o responsável pela exibição da aprendizagem em formato de "U", um dos grandes sucessos atribuídos ao modelo. Na aquisição natural da linguagem, as crianças apresentam esse desenvolvimento, mas são expostas, desde o primeiro momento, a todos os tipos de verbos, não apenas aos irregulares;
2) A utilização de Wickelfeatures pode gerar ambigüidades. Essa representação utiliza traços dos fonemas, mas não os segmentos propriamente ditos, o que pode causar equívoco na identificação de diferentes palavras. Outra falha da "Wickelfonologia" é o fato de ela não considerar a posição dos segmentos, mesmo sendo essa uma noção fundamental para que determinadas sub-regularidades dentro dos verbos irregulares sejam abstraídas, como, por exemplo, "ring"/"rang", "sing"/"sang";
3) O modelo baseia-se somente em informações fonológicas para produzir formas de passado, no entanto os falantes lançam mão de informações de outras fontes, não só a fonética, para flexionarem os verbos. Existem formas homófonas, como "ring"/"rang" e "wring"/"wrung", que dependem de diferenciação semântica para que as formas corretas de passado sejam geradas. Por causa disso, informação puramente fonológica não é o bastante para a produção de formas de passado.

A partir das críticas supracitadas, Pinker e Prince (1988, 1994) defendem a necessidade de regras para produzir as formas regulares da língua, por isso propõem um novo modelo para a estocagem de informações no léxico mental. Esse modelo é chamado "Teoria de Palavras e Regras" e será discutido na próxima subseção.

Após as críticas feitas ao modelo de Rumelhart e McClelland (1986), alguns outros modelos conexionistas foram elaborados, com o objetivo de sanar os problemas apresentados nessa primeira proposta de rede neural. Explicaremos, nos parágrafos subseqüentes, os modelos conexionistas mais relevantes já propostos na literatura lingüística e os sucessos por eles alcançados.

Plunkett e Marchman (1991) criaram um léxico artificial que preservava informações sobre os verbos do inglês. Esse léxico apresentava seqüências de CVC[42], VCC e CCV, que não correspondiam a conjuntos reais de sílabas do inglês, mas respeitavam a fonotática da língua. Informações sobre a posição dos segmentos eram, também, preservadas, ao contrário do que acontecia no modelo de Rumelhart e McClelland (1986). A noção de Wickelfeatures foi, portanto, abandonada. A freqüência de tipo (de um determinado padrão de forma de passado) e a de ocorrência (número de vezes em que uma ocorrência específica foi apresentada à rede) foram, também, controladas. Um corpus de 500 verbos foi utilizado como treinamento para o modelo. Essa rede continha algoritmos e divisões intermediárias: em vez de somente duas camadas (de entrada e saída), o modelo continha, também, "camadas ocultas" ("hidden layers"). A vantagem dessa arquitetura é que as camadas de entrada e saída são predeterminadas pela codificação de segmentos adotada, mas a camada "oculta" não é definida *a priori* e pode seguir qualquer tipo de representação que facilite a aprendizagem.

O modelo de Plunkett e Marchman (1991) foi capaz de dominar os diferentes tipos de flexão regular e irregular a que foi exposto, além de apresentar as mesmas fases de desenvolvimento da criança quando aprende os verbos do inglês, sem que fossem utilizados somente verbos irregulares no início do treinamento. Para que a rede aprendesse as irregularidades, uma interação entre freqüência de tipo e de ocorrência foi necessária, nos seguintes termos: mapeamentos arbitrários entre a forma de presente e a de passado dos verbos (baixa freqüência de tipo) foram aprendidos de forma adequada quando tinham freqüência de ocorrência suficientemente alta.

Uma peculiaridade importante da flexão de passado dos verbos do inglês é a seguinte: existe a regra geral para verbos regulares (adição do morfema *–ed*), mas há alguns verbos que, apesar de irregulares, pertencem a padrões de sub-regularidades (Cf. PINKER; ULLMAN, 2002a), tais como mudança de vogal ("ring"/"rang", "sing"/"sang"), nenhuma mudança ("cut"/"cut", "hit"/"hit") ou mudança de /d/ para /t/ ("lend"/"lent", "send"/"sent"). A presença desses casos de sub-regularidades dentro dos verbos irregulares pode ser considerada, em princípio, um problema para que redes neurais aprendam a produzir as formas corretas de passado; no entanto, o modelo de Plunkett e Marchman (1991) apresentou resultados melhores quando esse tipo de sub-regularidade foi introduzido no corpus. Os autores afirmam que, se não houvesse esses pequenos grupos de

[42] "C" deve ser entendido como "consoante" e "V", como "vogal".

verbos irregulares flexionados através dos mesmos processos, as idiossincrasias dos verbos dificilmente seriam aprendidas. É, portanto, por causa da freqüência desses tipos específicos de flexão que os verbos irregulares são armazenados no léxico mental e preservados.

MacWhinney e Leinbach (1991) testaram como um modelo conexionista responderia ao uso de um corpus real do inglês. Os autores selecionaram os 2.062 radicais de verbos mais freqüentes do corpus de Francis e Kučera (1982), juntamente com as formas de passado, particípio passado, particípio presente e terceira pessoa do singular desses verbos. Essas flexões compuseram um total de 6.090 itens. A representação fonológica do modelo era baseada em estruturas de vogais e consoantes. Informações semânticas foram preservadas para que não houvesse ambigüidades com formas homófonas.

Após o treinamento, a rede gerou corretamente 100% das formas regulares e 91% das irregulares. Na fase inicial de treinamento, houve generalização dos verbos irregulares, o que é compatível com os estágios de aquisição da flexão verbal pelas crianças. O maior sucesso alcançado por este modelo foi a comprovação de que uma rede neural poderia gerar formas adequadas de passado se fosse treinada com um corpus real do inglês. Isso demonstra que um modelo conexionista é capaz, até certo ponto, de replicar a cognição humana.

Daugherty e Seidenberg (1994) propuseram um modelo conexionista no qual havia três camadas utilizando 120 unidades para codificar estruturas CCCVVCCC (maior unidade silábica dos verbos monossílabos do inglês) e cada segmento era representado através da ativação de 15 traços articulatórios binários, como anterior, labial, coronal, vozeado, etc. A arquitetura do modelo consistia de três camadas (entrada, oculta e saída) e todas as unidades dessas camadas eram ligadas entre si.

Os verbos utilizados por esse modelo também foram retirados do corpus Francis e Kučera (1982), considerando informações sobre freqüência (verbos mais freqüentes eram apresentados mais vezes à rede) e quantidades de sub-regularidades proporcionalmente compatíveis com aquelas apresentadas no corpus adotado. Respeitando essas proporções, o treinamento da rede utilizou 309 verbos regulares e 24 irregulares.

Durante o período em que o modelo foi treinado, todos os verbos regulares e 92% dos irregulares foram corretamente produzidos. Outro resultado interessante

apresentado pelo modelo foi chamado pelos autores de "regularidade inconsistente", que pode ser definida como:

> [...] patterns such as BAKE-BAKED and FLIT-FLITTED, which obey the rule but have inconsistent rhyming "neighbors" (SEIDENBERG, 1992). Thus, even though BAKE-BAKED is rule-governed, performance may be impaired because the model must also encode the neighbors MAKE-MADE and TAKE-TOOK, which have irregular past tenses." (DAUGHERTY; SEIDENBERG, 1994, p. 372)

Daugherty e Seidenberg (1994) observaram que o modelo conexionista por eles proposto demonstrou diferenças no desempenho em relação a: 1) Verbos irregulares, 2) Regulares inconsistentes (com vizinhos irregulares) e 3) Inteiramente regulares (sem vizinhos irregulares), da seguinte forma: para gerar verbos regulares inconsistentes (que têm vizinhos similares foneticamente, mas irregulares), a rede demonstrou mais dificuldade do que para gerar verbos inteiramente regulares e menos dificuldade do que para gerar verbos irregulares. Ou seja, o modelo demonstrou um *continuum* de probabilidade de erros, sendo que os verbos irregulares foram os que mais geraram resultados errados, e os verbos totalmente regulares os que mais alcançaram sucesso. Os verbos regulares inconsistentes mostraram um desempenho intermediário nessa escala. Esse resultado é uma importante conquista do modelo, visto que, conforme veremos na subseção seguinte, teorias como a de Pinker (1991) não conseguem justificar o fato de verbos inteiramente regulares e regulares inconsistentes apresentarem resultados diferentes no que concerne à formação de passado. Já um modelo como o de Daugherty e Seidenberg (1994) pode explicar o comportamento de tais verbos (regulares inconsistentes) porque trata itens regulares e irregulares através do mesmo processo (rede de conexões), então efeitos de vizinhança lexical são previsíveis e até desejáveis.

Outra hipótese que Daugherty e Seidenberg (1994) testaram foi se o modelo seria suscetível a efeitos de freqüência. Os resultados demonstraram que, para os verbos regulares, freqüência não foi um fator significativo, ou seja, o fato de verbos regulares serem freqüentes ou infreqüentes não interferiu na capacidade de aprendizagem e generalização da rede; por outro lado, para os verbos irregulares, freqüência foi determinante para que o modelo produzisse mais acertos quando os verbos irregulares eram freqüentes.

Um traço comum a todos os modelos conexionistas resenhados até agora nesta subseção é o fato de eles proporem simulações apenas para a aprendizagem da flexão

verbal de passado do inglês. Conforme afirmamos anteriormente, essa morfologia costuma ser utilizada como exemplo do funcionamento de regras simbólicas no léxico mental do falante. Por conta disso, um desafio conexionista era observar se redes neurais puramente associativas poderiam recuperar informações sobre a flexão de verbos do inglês sem a utilização de regras. Como vimos, vários modelos foram bem-sucedidos nessa tentativa. Existem, no entanto, redes neurais que abordam outras questões lingüísticas além da morfologia verbal do inglês. Exemplo disso é o trabalho de Elman, que comentaremos a seguir.

Elman (1990) discute a representação do tempo em redes neurais. O autor propõe que o tempo seja representado não como uma dimensão da entrada ("input"), mas a partir dos efeitos que ele tem no processamento de informações. A fim de testar como modelos conexionistas iriam processar a noção de tempo, o autor elabora e treina redes neurais para desempenhar tarefas lingüísticas específicas, como prever a ordem de segmentos em uma palavra e predeterminar a seqüência de palavras em uma oração. Essas são habilidades que dependem do tempo, já que, na produção da fala, as seqüências de segmentos e palavras são temporalmente condicionadas. Através de sucessivos treinamentos para que pudesse aprimorar o desempenho, a rede elaborada por Elman (1990) foi capaz de aprender a estrutura temporal de sons em palavras e de palavras em frases. Para que essa aprendizagem ocorresse, a rede operou por generalizações estatísticas a partir da co-ocorrência de estruturas em uma linha temporal (segmentos em itens léxicos e palavras em sentenças).

Elman (1998b) elabora uma rede treinada para lidar com sintaxe. Avaliando a questão da necessidade de experiência perceptual para que a aprendizagem ocorra, o autor apresenta o seguinte desafio para ser testado por uma rede:

> Can a network trained only on sentences in which a given noun (e.g. "boy") appears only in subject position, deal appropriately with novel sentences in which that word appears in another syntactic context, e.g. object position[?]. (ELMAN, 1998b, p. 07)

A fim de simular a capacidade de generalização de modelos conexionistas, uma rede neural foi treinada para processar seqüências de sentenças, com uma palavra de cada vez, e prever a palavra seguinte. As palavras adotadas nesse treinamento diferiam em freqüência de ocorrência e os verbos se distinguiam em relação a complementos possíveis e a complementos preferíveis (quantidade de vezes em que um verbo aparecia com o mesmo

complemento). Vários corpora de tamanhos diferentes (variando entre 20 e 5.000 sentenças) foram criados a fim de verificar o desempenho da rede nessa generalização. Nessas sentenças, a palavra "boy" apareceu algumas vezes como sujeito, mas nunca como objeto direto. Nessa posição, apareciam outros sujeitos animados, tais como "adult", "girl", "Sue", mas não "boy". Então, a intenção dessa simulação era verificar se a rede consideraria "boy" um objeto direto possível, mesmo que, no corpus utilizado para o treinamento, essa palavra nunca tivesse ocorrido nessa posição.

Os resultados demonstraram que, apesar dessa lacuna no corpus (o fato de "boy" nunca ter ocorrido como objeto direto), a rede considerou que essa seria uma palavra possível nessa posição sintática. A explicação do autor para esse resultado é a seguinte:

> [...] although "boy" is never seen in direct object position, it is seen in shared contexts with other human words. For example humans (but not other animals, food, etc.) appear as the agent of verbs such as "eat", "give", "transfer". Conversely, humans (including "boy") do not appear as agents of other verbs (e.g. "terrify", "chase", which in this language require animal agents). The word "boy" shares more in common with other human words than it does with non-human words, or with verbs.
> In networks, as for human, similarity is a powerful motive force which can drive generalization. (ELMAN, 1998b, p. 10)

Portanto, o fato de a palavra "boy" partilhar contextos com outros agentes humanos já foi o suficiente para que a rede pudesse operar em termos de generalização, fazendo com que "boy" aparecesse, também, em contextos em que outros agentes humanos ocorriam. Assim, o conhecimento da rede sobre as possibilidades de posição sintática para o item "boy" foi determinado pela informação sobre aquilo que outras palavras similares poderiam fazer. Esse é um forte indício da capacidade de generalização e associação da rede.

Outra discussão importante levantada pelo autor a partir da análise dessa simulação é a relação entre o crescimento do léxico e a emergência da gramática, no sentido de que léxico e gramática estão mutuamente interligados, de modo que saber uma palavra é saber como usá-la. Para que generalizações ocorram, não é preciso que haja experiência exaustiva com uma palavra em todos os seus contextos possíveis de uso. Desde que haja experiência suficiente com outras palavras similares, de modo a formar uma categoria, a rede será capaz de generalizar uma determinada palavra para novos usos. Um requisito indispensável para que generalizações dessa natureza ocorram é que o corpus a que tanto a

rede quanto o ser humano tenham acesso seja em número suficientemente grande para que associações sejam feitas. Por causa disso, modelos conexionistas, segundo Elman (1998b, p. 13-14), demonstram melhor desempenho quando são treinados com corpora maiores. Concluímos, portanto, que a gramática emerge de associações e generalizações sobre padrões existentes no léxico. A separação entre léxico e gramática é, conseqüentemente, insustentável em um modelo dessa natureza.

Hare e Elman (1995) analisam o percurso histórico da flexão verbal do inglês e elaboram um modelo conexionista com o objetivo de simular a trajetória histórica dessa flexão, analisando os seguintes pontos: i) Os fatores que levam à mudança ou à relativa imunidade das formas; ii) A existência de um simples mecanismo (conexionista) ou processamento dual (regras e memorização) para processar a flexão dos verbos do inglês.

Hare e Elman (1995) explicam que, no inglês antigo ("Old English"), os verbos, em termos de flexão, eram organizados em dois grandes grupos: fracos ("weak") e fortes ("strong"). Esses grupos correspondem, respectivamente, ao que atualmente são os verbos regulares e irregulares. Estatisticamente, a classe dos verbos fracos era maior que a dos verbos fortes. Com a evolução do inglês antigo para o inglês moderno ("Modern English"), observa-se que houve vários verbos irregulares que migraram para a classe dos regulares.

Com base em uma análise etimológica, os autores afirmam que, devido ao fato de a classe de verbos regulares ser maior desde o inglês antigo, esse padrão poderia ter se espalhado por todo o sistema verbal da língua, eliminando as flexões irregulares. No entanto, dois fatores contribuíram para que isso não ocorresse: i) A classe dos verbos regulares tinha alta freqüência de tipo, mas os verbos, individualmente, tinham baixa freqüência de ocorrência; por outro lado, o grupo dos verbos irregulares tinha baixa freqüência de tipo, mas alguns verbos apresentavam, isoladamente, alta freqüência de ocorrência; ii) Os verbos irregulares organizavam-se em subgrupos de similaridades fonológicas, o que tornava essas pequenas classes mais consistentes no léxico mental. Assim, na evolução histórica, os verbos irregulares tenderam a se regularizar, adotando a desinência *–ed*; no entanto, devido aos fatores supracitados, se um verbo irregular apresentava alta freqüência de ocorrência ou pertencia a uma classe de itens fonologicamente similares, esse nivelamento não aconteceu, porque a alta freqüência (de tipo ou de ocorrência) garantiu a manutenção do padrão irregular de flexão. Dentro das

subclasses irregulares, os verbos que, individualmente, apresentavam baixa freqüência de ocorrência também tenderam a se regularizar, ainda que sua classe sobrevivesse.

A partir dessa análise histórica da morfologia verbal do inglês, Hare e Elman (1995) elaboram um modelo conexionista a fim de observar se esse percurso evolutivo dos verbos poderia ser replicado por uma rede neural. Utilizando, basicamente, a mesma arquitetura adotada anteriormente por outros modelos conexionistas, os autores treinaram a rede com verbos do inglês antigo, considerando os dois grupos básicos de flexão (regular e irregular) e as freqüências de cada uma das subclasses irregulares. Os resultados apresentados por esse modelo conexionista indicam que: i) As classes que continham mais verbos foram aprendidas com mais probabilidades de acerto; ii) As classes menores apresentaram maiores índices de erros e, em geral, os verbos dessas classes tenderam a migrar para os paradigmas mais freqüentes; iii) Quando os verbos não apresentavam nem alta freqüência de ocorrência nem pertenciam a classes numerosas, tenderam a sofrer nivelamento analógico nas bases do padrão mais freqüente (regular); iv) Alguns verbos regulares adotaram flexão irregular, mas só quando partilhavam alguma similaridade fonológica com a classe de verbos irregulares que os atraía. Portanto, a rede neural pôde simular o desenvolvimento histórico dos verbos do inglês, apresentando percurso evolutivo similar ao que o sistema verbal sofreu ao longo dos tempos. A conclusão de Hare e Elman (1995) é de que a interação de três fatores distintos determinou essa evolução:

> Our claim has been that stability in the inflectional system results from the interaction of type frequencies (i.e., class size), exploitable regularities of classes (i.e., phonological consistency), and the token frequencies of individual verbs. Verbs which either belong to small classes, lack consistent defining characteristics, or are low in frequency should change most rapidly; change in other verbs will depend on the precise extent to which they posses the characteristics which make them resistant to assimilation. (HARE; ELMAN, 1995, p. 90)

Portanto, segundo os autores, a imunidade ou a suscetibilidade à mudança está relacionada à interação entre freqüência de tipo, freqüência de ocorrência e existência de similaridades fonológicas que definam os itens lexicais em termos de classes. Sendo assim, "[...] on our account analogical change in the network is *not* strictly based on type frequency, but instead on frequency as it interacts with phonological consistency." (HARE; ELMAN, 1995, p. 93). A simulação proposta por Hare e Elman (1995) comprovou que a analogia afetou igualmente as classes de verbos regulares e irregulares, o que sugere que os

dois tipos de flexão são processados por um simples mecanismo, ou seja, através de redes associativas, e não de regras.

Elman (2004) mostra um ponto de vista alternativo para o léxico mental, em que as palavras devem ser pensadas como estímulos que atuam nos estados mentais. A fim de testar essa proposta, o autor elabora uma rede neural que, através do contexto, é capaz de prever a palavra seguinte em uma frase. Após sucessivos treinamentos, a rede demonstrou capacidade de aprender propriedades de cada palavra, usando essas características para prever os possíveis sucessores de um item léxico em uma sentença.

Em linhas gerais, através dessa resenha sobre modelos conexionistas, percebemos que tal teoria forneceu grandes contribuições para o entendimento de como a mente humana armazena e acessa informações lingüísticas. Seidenberg (1997) enfatiza que o desenvolvimento da tecnologia e a organização de corpora de línguas diversas possibilitaram pesquisas sobre padrões estatísticos e probabilísticos intrínsecos às estruturas lingüísticas. Utilizando esses recursos, os modelos conexionistas puderam comprovar que: i) As crianças, a partir do contato com seus cuidadores, adquirem, inconscientemente, informações estatísticas sobre a língua; ii) A tarefa das crianças é aprender a língua, não sua classificação gramatical; iii) Redes neurais são capazes de fazer generalizações sobre a aplicação de padrões, sem que evidência negativa seja fornecida; iv) O fato de haver subclasses de regularidades na língua (conforme explicamos anteriormente) facilita a tarefa da criança em operar com generalizações abstratas.

É claro que muito ainda há por ser feito no sentido de aprimorar o desenvolvimento das redes neurais. No entanto, o sucesso dos modelos conexionistas é inegável:

> Os neurônios do cérebro estão conectados de modo muito mais complicado e intrincado do que as unidades numa rede conexionista, mas, embora as redes conexionistas estejam longe de simular modelos perfeitos das conexões neurológicas do cérebro biológico, elas representam o que de mais próximo existe das bases biológicas da cognição. (CIELO, 1998, p. 47)

Elman (1998a, p. 16-17) cita algumas deficiências que modelos conexionistas ainda apresentam. São elas:

- No desenvolvimento natural do ser humano, é inegável o papel que as peculiaridades físicas apresentam. Nossa noção de espaço, por exemplo, é

determinada pelas características do nosso sistema visual. Portanto, a experiência humana depende das idiossincrasias físicas. Além disso, a experiência é sempre individual, no sentido de que a representação que alguém possui do mundo deriva de sua forma peculiar de vivenciar os estímulos. Redes neurais desconsideram essas limitações e possibilidades corporais, por isso as simulações de aprendizagem só conseguem replicar parcialmente a estrutura e o funcionamento da cognição humana;

- Organismos biológicos apresentam comportamentos variados de acordo com a situação em que se encontram. Não há organismos que tenham apenas uma função ou um padrão de conduta único. Animais, por exemplo, buscam comida, parceiros para o acasalamento ou abrigos para o frio. Redes neurais, em contraste, são arquitetadas para determinada função (aprender o sistema verbal do inglês, por exemplo) e, no decorrer de seu treinamento, não fazem nada além de lidar com o objeto de estudo para o qual foram construídas. Nessas redes não existe, portanto, a interferência de outros aspectos da cognição humana;
- Cognição é um fenômeno social. Os seres humanos aprendem padrões de comportamento a partir da convivência em sociedade. Modelos conexionistas são treinados para habilidades específicas aprendidas pela rede em isolado, o que nega o poder da convivência social no desenvolvimento de capacidades cognitivas.

Percebemos, portanto, que o treinamento de modelos conexionistas desconsidera algumas características importantes do desenvolvimento e organização natural da cognição humana. No entanto, apesar das críticas acima, é inegável o sucesso que redes neurais alcançaram na tarefa de replicar a forma como o ser humano aprende e generaliza habilidades específicas.

Com relação a nosso objeto de estudo (grupos de plurais do PB terminados em *–ão*, *–l* e ditongo em *–u*), poderíamos supor que o modelo conexionista trabalharia com o conceito de auto-recorrência, ou seja, aprendizagem por grau de saliência para recalibragem da rede a partir de padrões de freqüência. Sendo assim, para o grupo de itens terminados em *–ão* no singular, supor-se-ia que, após sucessivos treinamentos, a rede seria capaz de aprender a freqüência do padrão *–ões* para o plural, aplicando essa desinência à maior parte das palavras desse grupo. Quanto aos itens pluralizados em *–ãos* e *–ães*, a rede preservaria essas desinências se a freqüência de ocorrência dessas palavras individualmente fosse

suficientemente alta para manter os tipos infreqüentes (Cf. HARE; ELMAN, 1995). Quanto às palavras terminadas em *–l* e ditongo em *–u*, o modelo conexionista apresentaria resultados semelhantes, visto que, após algumas rodadas de treinamento, a rede poderia depreender a freqüência da pluralização em *–is* (já que a distinção fonética entre *–l* e *–u* em fim de sílaba foi perdida) e aplicar tal morfema aos itens terminados em *–l* e ditongo em *–u*. Da mesma forma, a pluralização em ditongo em *–u* (tipo infreqüente) seria preservada em itens de alta freqüência de ocorrência. Observa-se, portanto, que a rede poderia proceder a um auto-ensinamento a partir da exposição aos padrões do léxico.

Conforme já afirmamos anteriormente, o Modelo Conexionista não será adotado formalmente em nossa análise de dados; no entanto, como essa proposta é pioneira em concepções sobre o léxico mental, consideramos que seria interessante comentá-la, a fim de que o leitor desta tese pudesse perceber, de forma mais ampla, o contexto geral dos diferentes modelos de léxico mental já propostos na literatura lingüística.

Finalizadas as discussões sobre modelos conexionistas, passaremos para a próxima subseção, na qual analisamos a Teoria de Palavras e Regras, proposta por Pinker e Prince (1988 e 1994).

4.2 Teoria de Palavras e Regras

Conforme mencionamos na subseção anterior, Pinker e Prince (1988) fazem várias críticas ao Modelo Conexionista de Rumelhart e McClelland (1986) e apresentam outra teoria sobre o léxico mental, propondo que haja dois mecanismos diferentes de acesso a itens regulares e irregulares. Utilizando a flexão verbal do inglês como exemplo, essa primeira versão da teoria de Pinker e Prince (1988) defende que formas regulares, como "walk"/"walk<u>ed</u>", "learn"/"learn<u>ed</u>", são geradas por regras simbólicas, e formas irregulares, como "eat"/"ate", "drink"/"drunk", são estocadas no léxico mental. A regra se aplicaria à maior parte dos verbos, além de ser, também, utilizada em generalizações e formação de neologismos. Por outro lado, formas irregulares, por serem flexionadas com uma grande variedade de modificações no radical (ou até da supressão desse, como

"be"/"was", "go"/"went"), não poderiam ser enquadradas em uma única regra, por isso precisariam ser estocadas na memória dos falantes.

Assim como o modelo de Rumelhart e McClelland (1986), essa teoria de Pinker e Prince (1988) também sofreu algumas críticas, a saber:

1) Em geral, é comum a tendência de adultos e crianças produzirem flexões verbais que regularizam verbos irregulares, como "come"/"comed", "eat"/"eated". No entanto, existem, também, casos em que os falantes "irregularizam" verbos que são regulares, como "drag"/"drug" ou "squeeze"/"squoze", cujas formas padronizadas seriam "drag"/"dragged" e "squeeze"/"squeezed" (regulares, portanto). Se existem dois mecanismos separados para lidar com as flexões verbais do inglês (memorização para verbos irregulares e regras para regulares), a interferência de flexões irregulares em itens regulares não deveria ocorrer. No entanto, isso se verifica na aquisição da linguagem e em erros cometidos por adultos;
2) Entre os verbos irregulares, existem classes de itens que se agrupam por similaridade fonológica entre o radical e a forma de passado. Exemplos desses casos são "ring"/"rang" e "sing"/"sang", "fling"/"flung" e "cling"/"clung", "grow"/"grew" e "blow"/"blew". Se apenas formas regulares são geradas por regras, como explicar que existam esses grupos de verbos irregulares cujos membros partilham similaridades fonológicas no radical e na forma de passado? A ocorrência de tais semelhanças parece indicar que há esquemas locais para a flexão de itens que têm propriedades similares. Além disso, a existência de tais classes (de itens com semelhanças fonológicas) faz com que verbos criados para finalidades experimentais sejam flexionados a partir do paradigma de sua classe. Um exemplo disso é o verbo "spling" (não existente), que foi flexionado como "splung" em experimento desenvolvido por Bybee e Moder (1983). Nesse caso, como a palavra é inventada e não está estocada no léxico, a regra geral ("default") de acréscimo de *–ed* deveria se aplicar. No entanto, esse verbo foi flexionado como irregular devido à comparação com "fling"/"flung", "cling"/"clung". Esse caso demonstra que as sub-regularidades dos verbos do inglês podem influenciar os falantes na escolha de uma forma de passado. Então, regras e memória associativa não atuam de forma tão independente na flexão de passado do inglês.

Pinker (1991) e Pinker e Prince (1994) admitem essas falhas da proposta teórica de 1988, reconhecendo que: i) A memorização deve ser restrita a itens que não têm relação entre o singular e o plural, como "be"/ "was" ou "go"/"went", já que, nos demais verbos, parte do conteúdo fonológico do radical é preservado na forma de passado; ii) Uma teoria baseada em regras e memorização não pode explicar a organização de verbos irregulares em classes fonologicamente similares e a conseqüente extensão desses padrões irregulares a verbos regulares. Após assumir essas críticas, Pinker (1991) propõe alguns ajustes na teoria:

> These problems [above-mentioned] call for a theory of language with both a computational component, containing specific kinds of rules and representations, and an associative memory system, with certain properties of connectionist models. In such a theory, regular past-tense forms are computed by a rule that concatenates an affix with a variable standing for the stem. Irregulars are memorized pairs of words, but the linkages between the pair members are stored in an associative memory structure fostering some generalization by analogy. Although *string* and *strung* are represented as separate, linked words, the mental representation of the pair overlaps in part with similar forms like *sling* and *bring*, so that the learning of *slung* is easier and extensions like *brung* can occur as the result of noise or decay in the parts of the representation that code the identity of the lexical entry. (PINKER, 1991, p. 475)

Admitindo que tanto teorias conexionistas quanto modelos baseados em regras têm razão em diferentes aspectos, Pinker (1999, p. 117-118) afirma o seguinte:

> Prince and I have proposed a hybrid in which Chomsky and Halle are basically right about regular inflection and Rumelhart and McClelland are basically right about irregular inflection. Our proposal is simply the traditional words-and-rules theory with a twist. Regular verbs are computed by a rule that combines a symbol for a verb stem with a symbol for the suffix. Irregular verbs are pairs of words retrieved from the mental dictionary, a part of memory. Here is the twist: Memory is not a list of unrelated slots […], but is associative, a bit like the Rumelhart-McClelland pattern associator memory. Not only are words linked to words, but bits of words are linked to bits of words.

Então, o que Pinker (1999) propõe é uma fusão de diferentes aspectos da Teoria Gerativa e do Modelo Conexionista. Para a Teoria de Palavras e Regras[43] ("Words-and-Rules Theory" ou "Dual-Processing Model" ou "Modelo de Regras"), as palavras regulares são derivadas por regras, o que é uma contribuição da Teoria Gerativa, e as flexões irregulares são estocadas individualmente no léxico, em uma memória associativa,

[43] O título adotado por Pinker (1999) para seu modelo teórico é "Words-and-Rules Theory", no entanto preferimos traduzir esse título para "Teoria de Palavras e Regras", a fim de manter a prevalência da língua portuguesa em nossa tese.

segundo propunha o Modelo Conexionista. A justificativa para essa fusão de diferentes teorias é o fato de que itens irregulares dependem da memória para serem relembrados, mas flexões regulares são previsíveis e podem ser aplicadas a quaisquer palavras, sem que essas sejam, necessariamente, memorizadas como itens individuais. Sendo assim, a Teoria de Palavras e Regras defende que ambas as abordagens (Gerativa e Conexionista) estão corretas, mas com relação a itens específicos (regulares e irregulares, respectivamente), não às palavras como um todo.

Por outro lado, segundo o ponto de vista de Pinker (1999), tanto o Gerativismo quanto o Conexionismo têm problemas. Em primeiro lugar, o Gerativismo, porque, ao postular regras para todas as flexões, precisa recorrer a estruturas profundas "implausíveis" (PINKER, 1999, p. 121) e não consegue encontrar justificativas para o fato de que até os verbos irregulares do inglês apresentam similaridades em suas formas de passado. Por sua vez, o Conexionismo também apresenta problemas, porque, ao negar a atuação de regras, não pode explicar por que todos os verbos regulares do inglês apresentam a mesma desinência (*–ed*).

A Teoria de Palavras e Regras, conforme originalmente proposta por Pinker e Prince (1988) e a reformulada (PINKER, 1991; PINKER; PRINCE, 1994) adotam o mesmo princípio básico de processamento dual: regras para itens regulares e memorização para itens irregulares. Em seu cerne, a teoria permanece a mesma, com pequenas alterações. Daugherty e Seidenberg (1994, p. 364) sistematizam as duas principais diferenças entre essa teoria reformulada e a teoria original, conforme proposta em Pinker e Prince (1988):

- Primeiramente, a memória para itens irregulares é substituída por uma memória associativa que relaciona os radicais irregulares e suas formas de passado. Com essa modificação, a teoria apresenta explicação plausível para as semelhanças entre formas irregulares de presente e passado, além de justificar a existência de "famílias de similaridades" nas flexões verbais;
- A segunda grande modificação da teoria é que ela passa a assumir que as generalizações são produzidas por uma rede, não por regras. Isso explicaria o fato de falantes "irregularizarem" verbos que são regulares, a partir de similaridades com uma rede formada por itens regulares que partilham traços fonológicos com verbos irregulares.

A teoria proposta por Pinker (1991) e Pinker e Prince (1994) é, basicamente, a mesma que os autores defendem até hoje em seus artigos mais recentes (PINKER; ULLMAN, 2002a, 2002b). Um dos pontos principais dessa corrente teórica é a distinção entre léxico e gramática, nos seguintes termos:

> The lexicon is a subdivision of memory containing (among other things) the thousands of arbitrary sound-meaning pairings that underlie the morphemes and simple words of a language. The grammar is a system of productive, combinatorial operations that assemble morphemes and simple words into complex words, phrases and sentences. (PINKER; ULLMAN, 2002a, p. 456)

Estocados no léxico estão, portanto, os fonemas e os morfemas que, combinados a partir das regras da gramática, formam as palavras da língua. A partir dessa divisão entre léxico e gramática, Pinker (1999, p. 17) afirma que a armazenagem lexical atua a partir de dois "caminhos" distintos: palavras e regras, acessadas da seguinte maneira:

> If a word can provide its own past tense from memory, the rule is blocked; elsewhere (by default), the rule applies. […] as long as a verb does not have a form in memory, the rule may be applied. The ability of a rule to apply elsewhere or by default – that is, to any word that does not already have a specified form in memory – is the source of its power.

Portanto, na produção de uma palavra, o primeiro caminho seria o do léxico, a fim de verificar se a forma procurada está estocada na memória; se ela não estiver memorizada, a regra se aplica e a palavra é regularmente flexionada. O esquema abaixo ilustra essas duas rotas de acesso às palavras (memorização e regra):

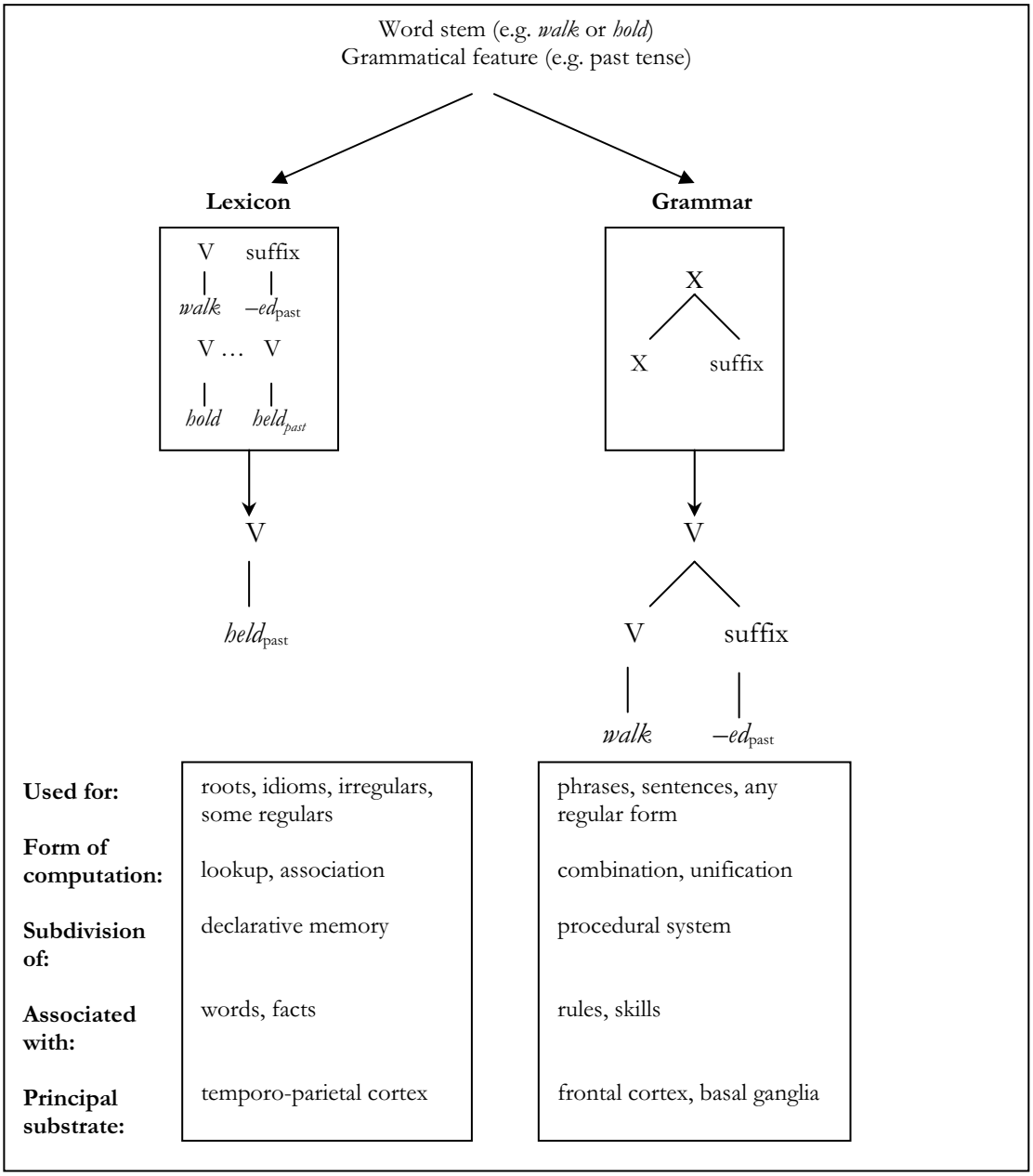


Figura 2: Ilustração sobre a Teoria de Palavra e Regras

Fonte: Pinker e Ullman 2002a, p. 457

A definição de Pinker (1999, p. 07) para a gramática de uma língua é a seguinte: "Grammar is an example of a combinatorial system, in which a small inventory of elements can be assembled by rules into an immense set of distinct objects." Essa gramática é dividida em diferentes módulos:

> Sentences are put together on an assembly line composed of mental modules, [...]. One is a storehouse of memorized words, the mental lexicon. Another is a team of rules that combine words and parts of words into bigger words, a component called *morphology*. A third is a team of rules that combine words into phrases and sentences, a component called *syntax*. (PINKER, 1999, p. 22)

Além de propor esses níveis específicos utilizados para a formação de palavras e sentenças, a Teoria de Palavras e Regras explica, também, a forma como esses diferentes módulos atuam:

> The three components pass messages about meaning back and forth with the rest of the mind so that the words correspond to what the speaker wants to say. This interface between language and mind is called *semantics*. Finally, the assembled words, phrases, and sentences are massaged by a set of rules into a sound pattern that we can pronounce when speaking or extract from the stream of noise when listening. This interface between language and the mouth and ear is called *phonology*. (PINKER, 1999, p. 22)

O diagrama a seguir ilustra a interação léxico mental e gramática, em sua estrutura e funcionamento, segundo a concepção de Pinker (1999):

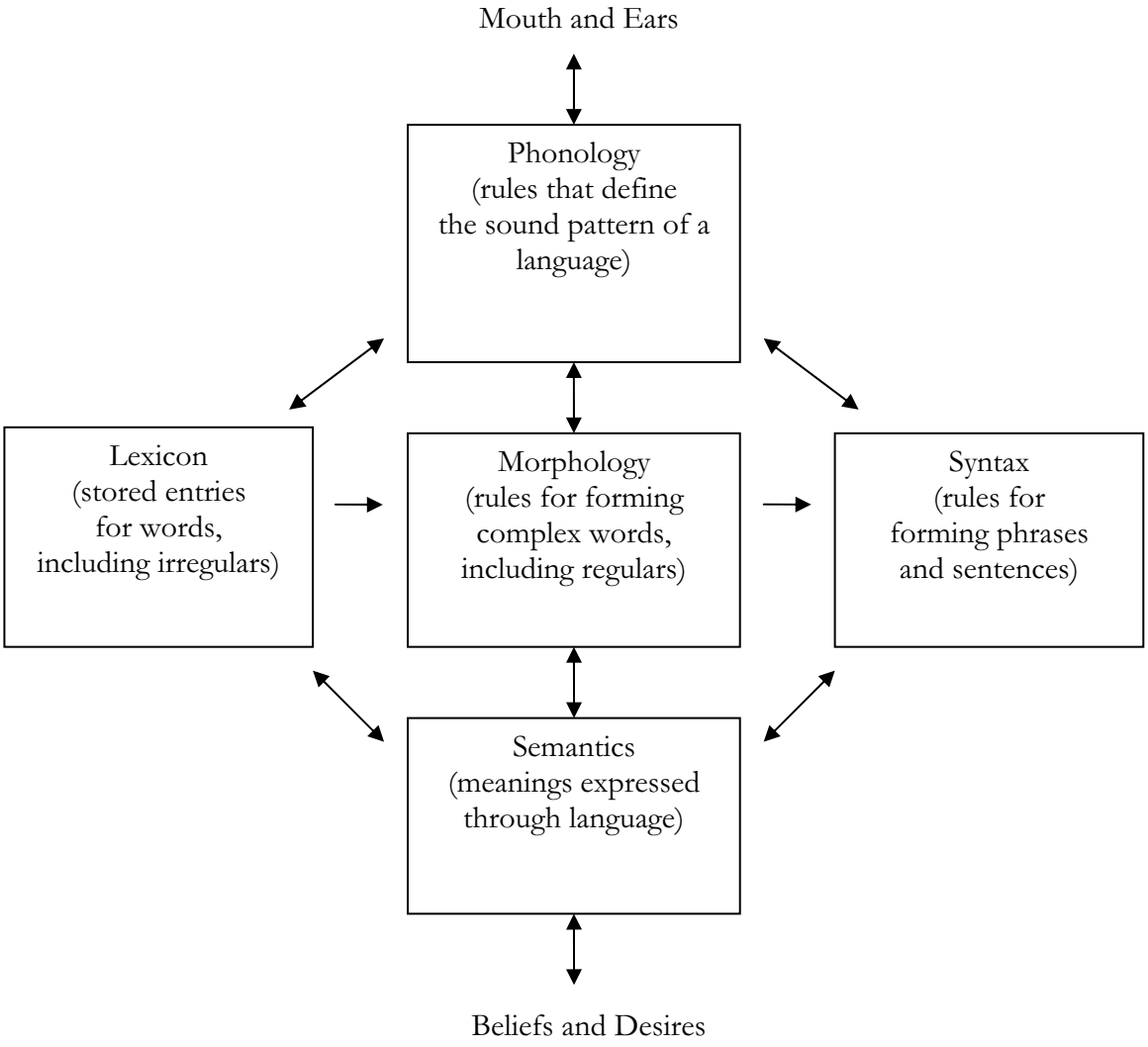


Figura 3: Diagrama sobre a estrutura e o funcionamento da gramática e do léxico mental

Fonte: Pinker 1999, p. 23

Conforme se pode perceber pelo diagrama exposto acima, a interação léxico e gramática proposta pela Teoria de Palavras e Regras é relativamente mecanicista e apriorística, uma vez que parte do princípio de que a linguagem serve para expressar crenças e desejos e que há um caminho inequívoco entre o pensamento e a boca/ouvidos dos interlocutores, passando pelos diversos componentes da gramática.

Para a Teoria de Palavras e Regras, as palavras são estocadas na mente do falante em suas formas primitivas. Além disso, o léxico mental contém, também, informações sobre desinências específicas (passado e plural, por exemplo) e regras sobre como essas podem ser combinadas aos radicais, a fim de formar palavras. Os radicais no léxico mental funcionam como entradas em um dicionário real: eles especificam a que

classe gramatical a palavra pertence, definem o significado da palavra e especificam seu som. Na Teoria de Palavras e Regras, existe, também, o conceito de "economia" das regras lingüísticas: "*One* past tense suffix is stored in the lexicon, […] and a separate module fiddles with its pronunciation: the rules of phonology, which define the sound pattern or accent of a language." (PINKER, 1999, p. 33). Portanto, apenas informações não--redundantes são estocadas e os itens léxicos são gerados a partir da aplicação de regras aos fonemas e morfemas armazenados.

Segundo Pinker (1999), o poder expressivo das regras para formar frases e/ou palavras pode ser assim formalizado:

- **As regras são produtivas:** especificando tipos de palavras e não palavras isoladamente, uma única regra pode ser aplicada a uma quantidade grande de itens. Além disso, as regras permitem, também, que quaisquer pensamentos possam ser expressos de forma criativa e inovadora, já que as regras são finitas, mas produzem um número infinito de sentenças e palavras possíveis dentro de uma língua;
- **As regras são simbólicas e abstratas:** assumem-se, por exemplo, estruturas determinadas para palavras e frases pertencentes a uma certa língua; no entanto, a regra não explicita os segmentos ou as palavras específicas que devem compor o item léxico ou a frase. Dentro do inventário fonético da língua e do léxico (dicionário), é possível gerar quaisquer palavras e frases a partir da aplicação de regras simbólicas e abstratas;
- **As regras são combinatórias:** embora haja uma grande possibilidade de palavras que podem ocupar determinado lugar na sentença, cada posição deve ser preenchida por classes específicas de palavras. Um predicado, por exemplo, requer verbo, ao passo que um sujeito requer substantivo ou pronome. Portanto, existem possibilidades específicas de itens léxicos que podem assumir determinados espaços dentro das sentenças de uma língua, ou seja, não é qualquer palavra que pode ocupar qualquer lugar dentro da frase.

Pinker (1991) e Pinker e Prince (1994) apresentam três tipos de argumentos que comprovariam que a Teoria de Palavras e Regras é o modelo mais adequado de léxico mental. Essas evidências baseiam-se em informações sobre freqüência, similaridade e dissociações neurológicas. Examinaremos mais detalhadamente cada um desses fatores.

Com relação a efeitos de freqüência, Pinker (1991) afirma que esses podem ser encontrados apenas na classe dos itens irregulares, já que eles estão estocados no léxico e, portanto, a sua representação na memória se torna mais forte quando são mais usados; por outro lado, itens regulares, como são derivados por regras e não se encontram no léxico, não devem sofrer efeitos de freqüência.

Argumentos de que freqüência afeta apenas os verbos irregulares são os seguintes: 1) Em experimentos psicolingüísticos, verbos irregulares de baixa freqüência são mais propensos a regularizações por crianças e adultos ("breaked" em vez de "broke", por exemplo); 2) O inglês possui dezenas de verbos irregulares, mas são justamente os mais freqüentes – "be", "do", "go", "have" – que apresentam as irregulares mais idiossincráticas. Nesse sentido, a alta freqüência de ocorrência desses itens ajudaria a manter o paradigma irregular; 3) Em experimentos realizados, Prasada, Pinker e Snyder (1990) *apud* Pinker (1999) analisaram a velocidade com que os falantes pronunciariam o passado de verbos regulares e irregulares de baixa e alta freqüência de ocorrência. Os autores observaram que verbos irregulares com baixa freqüência de ocorrência foram acessados no léxico mental mais devagar que verbos irregulares com alta freqüência de ocorrência; em contrapartida, freqüência de ocorrência não foi relevante para verbos regulares, visto que tanto os freqüentes quanto os infreqüentes foram acessados na mesma média de velocidade. Esses experimentos comprovariam que freqüência afeta somente palavras irregulares, já que apenas elas estão armazenadas no léxico mental e dependem da memória para que sejam recuperadas.

Quanto à similaridade entre radicais e formas de passado para verbos irregulares, Pinker (1991, p. 477) afirma que, já que a teoria prevê que a memória é de natureza associativa, é natural que verbos irregulares se agrupem em famílias de similaridades. De fato, crianças apresentam menor índice na regularização de verbos irregulares quando esses se encontram dentro de tais grupos. Há evidências, então, que essas similaridades fonológicas entre os verbos fortalecem a classe como um todo, fazendo com que tais itens sejam memorizados mais facilmente. Com relação à produção dos itens regulares, a teoria prevê que não deve haver nenhuma influência de outros verbos irregulares, já que a regra é o "default", aplicado a quaisquer verbos regulares, independentemente de sua forma fonológica. Verbos regulares podem ter similaridades fonológicas com verbos irregulares, mas, ainda assim, são flexionados a partir da regra: radicais como "need" e "bleed", "blink" e "shrink" são fonologicamente similares, mas isso

não faz com que a flexão dos irregulares interfira na dos regulares. Assim, tais verbos apresentam as flexões "need"/"needed" e "bleed"/"bled", "blink"/"blinked" e "shrink"/"shrank", preservando sua estrutura interna (regular e irregular, respectivamente, para cada par de palavras), apesar de seus radicais serem fonologicamente semelhantes. Não existe, portanto, interferência de verbos irregulares na formação de regulares, ainda que haja similaridades fonológicas. Segundo Pinker (1991) e Pinker e Prince (1994), esses são indícios de que flexões regulares e irregulares são instâncias diferentes de processamento, gerenciadas através de mecanismos diferentes (regras e palavras, respectivamente).

Com relação às dissociações neurológicas, Pinker (1991) e Pinker e Prince (1994) comentam casos de falantes que sofreram danos em partes do cérebro (como afásicos) e ainda apresentam alguma habilidade para processamento lingüístico. Analisando a fala de tais indivíduos, os autores encontraram casos de perda gramatical, em que pessoas não conseguiam mais aplicar a regra de flexão aos verbos regulares, mas os irregulares eram perfeitamente empregados. Tal fato foi considerado pelos autores um argumento de que itens regulares e irregulares apresentam diferentes módulos na mente humana.

Apesar das diferentes formas de acesso a itens regulares e irregulares, a Teoria de Palavras e Regras não descarta, também, o fato de que, em casos excepcionais, itens regulares possam ser listados diretamente na memória, sem que seja necessário recorrer a regras para produzi-los:

> [Words-and-Rules Theory] does not posit that regular forms are *never* stored, only that they do not *have to* be […]. It would be difficult to prohibit regular forms from ever being stored, given that human memory can acquire many kinds of verbal material (e.g. idioms, clichés, poems). (PINKER; ULLMAN, 2002a, p. 458)

Portanto, algumas formas regulares podem ser armazenadas na memória, embora esse não seja o caminho natural para tais palavras. Casos de unidades automatizadas na fala, tais como expressões coloquiais comumente usadas ("chunks") podem ser estocadas lexicalmente, visto que os falantes, na maior parte das vezes, automatizam essas seqüências e não têm consciência das unidades que as formam. Pinker (1999) cita experimentos que demonstram que as pessoas demoram mais tempo para reconhecer itens regulares que são usados raramente. Se tal diferença de acesso a palavras regulares freqüentes e infreqüentes existe, é porque algumas formas, mesmo que sejam

regulares, podem ser guardadas diretamente na memória, sem que precisem ser derivadas por regras. A conclusão que se pode alcançar a partir dessas afirmações é o fato de que os falantes constroem uma memória para formas regulares muito usadas, mesmo que essa memória, teoricamente, não seja necessária. Nas palavras do autor: "Human memory is not a scarce resource reserved for the incompressible nuggets that cannot be generated by rules. If a word is common enough, we can look it up directly, rather than breaking it into parts and looking up the parts." (PINKER, 1999, p. 138). Então, conforme essa citação demonstra, a posição de Pinker (1999) difere da de Pinker (1991), porque, anteriormente, o autor negava que freqüência pudesse afetar os itens regulares; já em 1999, essa possibilidade é admitida, pelo menos de forma parcial, pelo autor. Observa-se, no entanto, que o processamento por palavras e regras tem prevalência e a memorização ocorre apenas para itens idiossincráticos. Pinker (1999) menciona alguns desses casos (em que ocorre memorização para itens regulares): a) Crianças, antes de abstraírem a regra para gerar formas regulares na língua, precisam memorizar todas as palavras a que têm acesso, inclusive as regulares; b) Se um verbo tem duas formas possíveis de passado, uma regular e outra irregular, a presença do regular é imprevisível, então esse deverá, também, estar estocado na memória do falante; c) Verbos regulares que rimam com um irregular também podem ser estocados na memória, porque um efeito analógico pode fazer com que um verbo regular seja flexionado como um irregular. Observamos, portanto, que, para a Teoria de Palavras e Regras, a memorização de itens regulares opera apenas em casos excepcionais, como os listados acima. De forma geral, prevalece a derivação por regras para palavras regulares.

No caso dos plurais sob investigação nesta tese, a Teoria de Palavras e Regras postularia a existência de uma regra geral ("default") aplicada para a formação de plural. No PB, esse seria o caso da adição do morfema *–s*, já que essa é a regra mais produtiva na língua e a que menos sofre restrições morfofonológicas para sua aplicação. Sendo assim, os grupos de plurais em *–ão* + *–s* (*–ãos*) e *–u* + *–s* (*–us*) seriam regulares, já que seguem o padrão default da língua. Por outro lado, itens pluralizados através dos morfemas *–ães*, *–ões* (para os casos de singular em *–ão*) e *–is* (para os casos de singular em *–l*) seriam considerados irregulares, já que sofrem processos morfofonológicos diferentes do padrão (pluralização em *–s*). Seguindo o raciocínio de que itens regulares são derivados por regras e itens irregulares estão estocados na memória, a Teoria de Palavras e Regras postularia que os itens pluralizados em *–ão* + *–s* (*–ãos*) e *–u* + *–s* (*–us*) seriam gerados por regras, ao passo

que os demais (*–ães*, *–ões* e *–is*, conforme exposto acima) seriam estocados na memória do falante. A partir dessas premissas básicas, poderíamos formular as seguintes hipóteses para o comportamento desses plurais, de acordo com os pressupostos da Teoria de Palavras e Regras:

- Se os falantes pluralizam as palavras através da adoção de uma regra simbólica, deverão adicionar um sufixo *–s* ao fim da palavra, já que esse é o "default" de pluralização no PB. Então, os plurais em *–ãos* (para itens terminados em *–ão* no singular) e em *–s* (para itens terminados em *–l* e *–u* no singular, devido à perda de distinção fonética) devem ser os preferidos pelos falantes;
- Se os falantes estocam no léxico apenas palavras irregulares, e as regulares são derivadas por regras, não deve haver efeitos de freqüência em palavras regulares (pluralizadas em *–ão* + *–s* ou *–u* + *–s*). Efeitos de freqüência são previstos somente para itens irregulares (pluralizados através de *–ões*, *–ães* e *–is*), porque estes estão listados no léxico mental e, portanto, dependem da memória para ser relembrados;
- Se os falantes pluralizam as palavras regulares através de regras, utilizarão a vogal final como o indício para adicionar o plural em *–s*. Já que todas as palavras em *–ão* e *–u* no singular terminam em vogal, é de se esperar que o plural em *–s* seja o mais freqüente.

Portanto, essas seriam as hipóteses principais da Teoria de Palavras e Regras para os grupos de plurais do PB considerados nesta tese. Na próxima subseção, trataremos do Modelo de Redes.

4.3 Modelo de Redes

O Modelo de Redes[44] ("Network Model" ou "Usage-Based Phonology" ou "Fonologia de Uso") foi proposto por Bybee (1985, 1995, 2001) e apresenta peculiaridades

[44] Na verdade, o Modelo de Redes não é, propriamente, um modelo formal, como o Modelo Conexionista, que propõe a elaboração de redes neurais para entender como a cognição opera. O uso do termo "modelo", quando aplicado ao Modelo de Redes, refere-se a uma teoria sobre a organização do léxico mental.

que o diferem das duas teorias resenhadas anteriormente nesta tese (Modelo Conexionista e Teoria de Palavras e Regras).

Bybee[45] (1985, 1995) discute a controvérsia existente entre diferentes modelos de léxico mental. Segundo ela, há, basicamente, duas propostas a esse respeito: i) A Teoria Gerativa (CHOMSKY; HALLE, 1968) propõe que somente os morfemas estariam estocados na memória e, sendo assim, todas as palavras derivadas seriam formadas por regras que combinassem tais morfemas para formar itens léxicos; ii) A teoria de Vennemann (1974) postula que todas as palavras, independentemente de seu tamanho e complexidade morfológica, estariam estocadas na memória. A existência de regras também é aceita nesse modelo, mas a estocagem de unidades inteiras no léxico é prioritária.

Segundo Bybee (1985, 1995), o modelo proposto por ela situa-se no meio desses dois extremos (estocagem por regras / armazenagem de todas as formas simples e complexas). A autora afirma que a existência de memória lexical associativa é inegável, já que formas irregulares precisam ser estocadas no léxico mental (BYBEE, 1985, p. 112). Evidência disso é o fato de que não existe mapeamento fonológico possível entre algumas formas supletivas, como "be"/"was" ou "go"/"went". Não há regra que consiga justificar a derivação de tais flexões a partir da forma primitiva desses verbos, portanto casos como esses requerem, indubitavelmente, estocagem lexical. Por outro lado, a existência de regras também é inquestionável, visto que os falantes lançam mão delas quando precisam flexionar formas que nunca ouviram antes.

Já que estocagem por associação e regras são duas alternativas admitidas pelo Modelo de Redes, alguém poderia argumentar que essa proposta de léxico mental é uma versão nova da Teoria de Palavras e Regras. No entanto, o Modelo de Redes propõe que itens regulares também podem ser estocados na memória lexical, ponto este contrário à Teoria de Palavras e Regras, que prevê que itens regulares são derivados por regras e não precisam, necessariamente, ser armazenados no léxico mental[46]. Bybee (1985) afirma que formas regulares podem ser listadas no léxico mental, e o fator que determina a estocagem ou não de tais itens é a freqüência de ocorrência: "[…] low-frequency items are analyzed

---

[45] Em seu livro de 1985, "Morphology: *a study of the relation between meaning and form*", Bybee ainda não chega a cunhar de forma sistemática o modelo de léxico mental elaborado por ela. Nesse livro, são apresentadas idéias gerais sobre morfologia e as bases teóricas do Modelo de Redes são esboçadas, mas o próprio termo "Network Model" não é mencionado. É apenas em 1995 que essa nomenclatura é adotada.

[46] Conforme vimos na subseção anterior, para a Teoria de Palavras e Regras, itens regulares podem ser armazenados no léxico, mas apenas em casos específicos, como os que mencionamos naquela subseção. De um modo geral, prevalece a noção de regras para itens regulares e memória associativa para os irregulares.

and understood in terms of other items, while high-frequency words, complex or not, may be autonomous, and processed unanalyzed." (BYBEE, 1985, p. 124). Portanto, o Modelo de Redes afirma que itens regulares podem ser estocados desde que apresentem alta freqüência de ocorrência na língua. O fato de serem muito usados garante a tais itens uma autonomia lexical, permitindo que sejam listados na memória do falante, apesar de poderem, também, ser derivados por regras. É, portanto, a repetição de um item lingüístico que avaliza sua armazenagem no léxico mental. No caso da Teoria de Palavras e Regras, a estocagem das palavras é determinada, primordialmente, por sua estrutura, não pelo seu uso.

Outro ponto divergente entre essas duas teorias é a distinção entre léxico e gramática. Na Teoria de Palavras e Regras, léxico e gramática são dois módulos separados, visto que o léxico armazena itens irregulares e a gramática flexiona os regulares; para o Modelo de Redes, essa separação formal não existe, porque as palavras já são armazenadas em suas formas flexionadas, independentemente de sua (ir)regularidade. A princípio, esses são os pontos básicos que diferem o Modelo de Redes e a Teoria de Palavras e Regras. Ao final desta subseção, vamos aprofundar as diferenças entre esses dois modelos de léxico mental.

Ao longo do percurso evolutivo da teoria proposta inicialmente por Bybee (1985), observamos que a autora, inicialmente, cogita a possibilidade de estocagem por palavras e regras; no entanto, em seus trabalhos mais recentes, embora a autora não negue explicitamente a existência de regras, ela enfatiza que palavras inteiras são listadas no léxico mental, o que, teoricamente, negaria a atuação de regras simbólicas (BYBEE, 2001).

Bybee (1985, p. 134) apresenta as propriedades fundamentais do modelo proposto por ela:

1. As palavras são representadas no léxico a partir de diferentes "forças lexicais" ("lexical strength"). A cada vez que uma palavra é acessada, sua representação lexical torna-se mais forte;
2. Palavras regulares derivadas são listadas no léxico desde que tenham alta freqüência de ocorrência. Tais palavras formam conexões lexicais com outros itens similares em termos fonológicos e semânticos. Da mesma forma, itens irregulares são, também, estocados no léxico mental;

3. Palavras regulares derivadas que não têm alta freqüência de ocorrência são mapeadas a partir de representações de itens com que partilham as mesmas propriedades semânticas e fonológicas;
4. As conexões lexicais são mais ou menos fortes a partir do número de traços que as palavras constituintes dessas conexões partilham entre si.

No Modelo de Redes, a unidade básica de estocagem lexical é a palavra inteira ou até mesmo expressões freqüentemente utilizadas:

> [...] I am abandoning the idea that the items in the lexicon are all either morphemes or words, and accepting the idea that units of different sizes may be represented. Except for some very productive affixes, there is little evidence for sequences any smaller than the word, and there is evidence of various sorts that the word is a unit with psychological reality. For instance, the word is the relevant unit for lexical diffusion of sound change. (BYBEE, 1985, p. 135)

Como evidências para o fato de que, em geral, a palavra é a unidade fundamental de armazenamento, podemos citar dois fatos: 1) É somente a partir da comparação de palavras inteiras dentro de um paradigma que o falante depreende a estrutura interna dos itens lexicais; 2) Expressões freqüentemente usadas são mais propensas a fenômenos de redução fonético-fonológicos em geral. Bybee e Scheibman (1999), analisando a redução da partícula "don't", no inglês americano, observaram que tal processo ocorria mais na expressão "I don't know", que era a mais freqüente do corpus analisado. Esse seria um exemplo de que expressões altamente freqüentes são processadas como unidades inteiras, o que faz com que sejam suscetíveis a fenômenos de redução verificáveis somente nesses contextos e ausentes (ou existindo em menor amplitude) em contextos alternantes.

Bybee (1995) refina alguns dos pressupostos teóricos que já haviam sido esboçados em Bybee (1985). Segundo a autora, o ponto diferencial do Modelo de Redes com relação a modelos propostos anteriormente é o seguinte:

> A major difference between this model, which I will call the *network* model, and structuralist models containing rules, is that actual usage in terms of both type and token frequency plays an important role in establishing and maintaining representation. (BYBEE, 1995, p. 428)

A freqüência de uso das estruturas lingüísticas é, portanto, de alta relevância para as representações mentais. O Modelo de Redes analisa a freqüência a partir de duas medidas diferentes. Conforme afirmamos no capítulo anterior, "Token or text frequency is the frequency of occurrence of a unit, usually a word or morpheme, in running text." […] "Type frequency refers to the dictionary frequency of a particular pattern, such as a stress pattern, an affix, etc" (BYBEE, 2002a, p. 604). Um exemplo de freqüência de ocorrência ("token") seria a quantidade de vezes que um determinado item ocorre em um corpus de uma língua específica; freqüência de tipo ("type") seria a contagem de quantas palavras no dicionário dessa língua apresentam um determinado padrão de plural ou sufixo, por exemplo[47].

Na análise de fenômenos lingüísticos, essas duas abordagens de freqüência interagem e suscitam efeitos diferenciados no léxico:

1) Palavras altamente freqüentes têm representação lexical mais forte e, por isso, sofrem menos análise e são menos dependentes de suas formas primitivas. Palavras infreqüentes, por outro lado, não têm autonomia lexical e, para serem ativadas, dependem de conexões com outros itens lexicais com que partilham propriedades semânticas e fonológicas;
2) Palavras freqüentes são acessadas mais rapidamente do que palavras infreqüentes, justamente porque sua representação lexical é mais forte e sua memória é ativada com mais rapidez;
3) A manutenção de uma palavra com flexão irregular depende de sua freqüência de ocorrência. Palavras irregulares e freqüentes têm representação lexical mais forte, portanto costumam resistir a mudanças analógicas[48]; por outro lado, palavras irregulares e infreqüentes não são suficientemente reforçadas no léxico mental para que sua irregularidade se mantenha, por isso é comum que adotem os paradigmas mais freqüentes da língua;

---

[47] Nesta tese, para garantir a coerência de nomenclatura, adotaremos as traduções "tipo" e "ocorrência" para os termos que Bybee (1995) nomeia como "type" e "token", respectivamente.

[48] Segundo Bybee (2001), palavras irregulares e de alta freqüência de ocorrência resistem mais a mudanças analógicas (adoção de um paradigma diferente do de sua classe, a partir do efeito da analogia). No entanto, a repetição dos itens léxicos gera automatização e sobreposição dos gestos articulatórios, o que faz com que itens altamente freqüentes sejam mais propensos a reduções fonéticas em geral (Cf. PAGLIUCA; MOWREY, 1987, e BROWMAN; GOLDSTEIN, 1992).

4) Palavras derivadas e de freqüência de ocorrência baixa ou média formam conexões mais fortes do que palavras derivadas e de alta freqüência de ocorrência. Isso ocorre porque os itens infreqüentes dependem de sua classe para ser relembrados, portanto, a cada vez que são acessados, reforçam a coesão de sua rede; por outro lado, itens derivados altamente freqüentes ganham autonomia lexical e sua ativação não é feita necessariamente através de sua classe, por isso o acesso a itens derivados e freqüentes não reforça a conexão da classe como um todo. Portanto, é a freqüência de tipo, não de ocorrência, que garante a produtividade de uma classe;
5) Como decorrência do item 4 acima, os padrões flexionais encontrados em palavras altamente freqüentes não tendem a se espalhar para outros itens lexicais, porque o acesso a tais itens (mais freqüentes) não reforça a rede; por outro lado, as flexões de itens com freqüência baixa e média podem, mais facilmente, ser adotadas por novas palavras, já que o acesso a palavras pouco freqüentes é importante para a força lexical da rede: "[...] if a class contains both high-frequency and medium-frequency items, it is the medium-frequency items that contribute most to the productivity of the class, and to the establishment of the parameters that define the class." (BYBEE, 1985, p. 134).

Com relação à organização do léxico mental, Bybee (1995) afirma o seguinte:

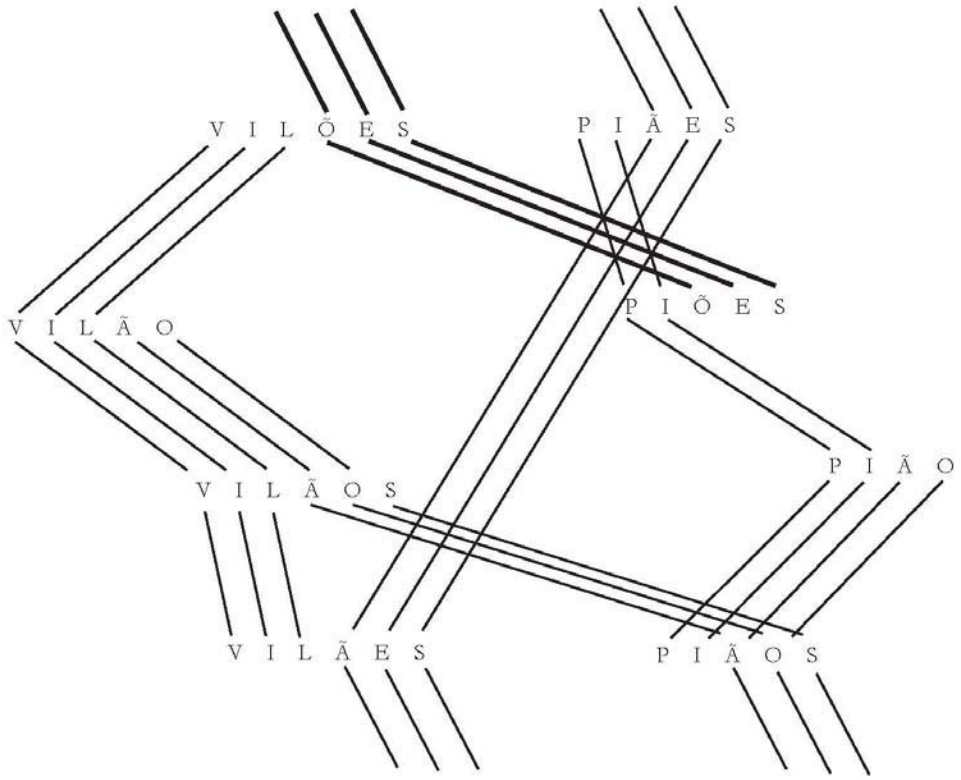
> Even though words entered in the lexicon are not broken up into their constituent morphemes, their morphological structure emerges from the connections they make with other words in the lexicon. Parallel sets of phonological and semantic connections, if they are repeated across multiple sets of words, constitute morphological relations […] (BYBEE, 1995, p. 428-429)

Portanto, as palavras estão listadas no léxico mental em forma de redes, sendo agrupadas de acordo com identidade ou similaridade fonológica ou semântica. A estrutura morfológica emerge dessas conexões, fazendo com que uma determinada classe paradigmática seja mais ou menos forte, dependendo da quantidade de itens que se flexionam através dela. Com relação à extensão dessas redes, Bybee (1995, p. 430) afirma o seguinte:

> Sets of words having similar patterns of semantic and phonological connections reinforce one another and create emergent generalizations describable as schemas. New items or items whose connections are not known or are weak can be fitted into these schemas. The likelihood of the schema being extended to

> new items is directly dependent upon two factors: (i) the defining properties of the schema and (ii) its strength, the latter property being derivable from the number of items that reinforce the schema. If the defining properties of the schema are very specific, the schema will be restricted in its application to new forms, and result in lower productivity. If the schema is very open, placing few restrictions on the items to which it can apply, its productivity is the strength of the schema, which is based directly on its type frequency – the higher the frequency of the pattern described in the schema, the greater are its chances of applying to new items. (BYBEE, 1995, p. 430)

Então, para que um esquema seja aplicado a novas palavras, dois fatores são determinantes: a quantidade de itens que se flexionam através dele e a existência ou não de restrições. Se um esquema tem muitas restrições fonológicas ou morfológicas, poucos itens poderão migrar para essa classe; por outro lado, se um esquema é relativamente aberto, não apresentando muitas restrições, mais receptivo ele será à entrada de novos itens. No caso dos plurais das palavras em *–ão*, por exemplo, poderíamos propor o seguinte esquema de conexões:



Esquema 1: Interconexões fonéticas para palavras terminadas em *–ão* no singular e no plural

No esquema acima, ilustrativo da organização do léxico mental segundo o Modelo de Redes, algumas peculiaridades merecem ser comentadas: 1) Os itens são armazenados em sua forma singular e plural e, como as duas palavras adotadas ("vilão" e "vulcão") podem apresentar mais de uma possibilidade de plural, as formas variantes são, também, estocadas, desde que o falante já as tenha ouvido; 2) Existem inter-relações fonológicas entre o singular e o plural e entre os itens plurais entre si, através das diferentes desinências de pluralização; 3) O acesso a um item plural reforça sua conexão com o singular, por isso o fato de haver mais itens em *–ão* pluralizados como *–ões* leva à generalização morfológica de que *–ões* é o plural mais produtivo para *–ão*. Essa generalização faz com que o plural *–ões* seja o mais forte no léxico mental (por isso a linha em negrito) e se aplique a novos itens da classe em *–ão*, sobretudo àqueles de baixa freqüência de ocorrência no plural, que não têm memória suficiente para que sejam relembrados de forma independente e precisam de sua classe para que sejam flexionados.

O Modelo de Redes sugere, portanto, que a armazenagem lexical é organizada em termos de freqüência de tipo e de ocorrência, considerando a similaridade fonológica e semântica entre os itens armazenados. Para que relações de semelhança entre os itens léxicos possam ser estabelecidas, a estocagem precisa ser redundante, de modo que diferentes pronúncias da mesma palavra, assim como diferentes contextos situacionais em que uma palavra foi pronunciada, estão presentes nas representações mentais de informações lingüísticas. Bybee (2001, p. 20-21) afirma que a armazenagem de informações redundantes é necessária para a categorização de novos exemplares e para a formação de protótipos. Para ilustrar essa questão, a autora fornece o exemplo de um cachorro. Existem inúmeras espécies de cachorros, e eles partilham entre si mais semelhanças que diferenças. Ainda assim, nossa memória não armazena apenas os traços não-redundantes deles. Todos os tipos de cachorros são memorizados, a fim de que comparações entre eles possam ser feitas e que suas propriedades intrínsecas possam ser deduzidas. O mesmo ocorre com as palavras: a estocagem de detalhes fonéticos é imprescindível para que categorizações e organizações lexicais sejam formadas. Nessa perspectiva, já que a gramática emerge do uso da língua, as mudanças sonoras são vistas como naturais e intrinsecamente ligadas à natureza mutável da língua:

> In the usage-based model being explored here, memory representations are categorizations of tokens of use. Thus, if the tokens of use begin to change, the center of the category will also gradually change. The direct representation provided by phonetic categorization allows for gradual phonetic change in categories, as well as in particular lexical units. (BYBEE, 2001, p. 58)

No Modelo de Redes, como há representações mentais múltiplas para um mesmo item, a variação sonora apresenta gradualidade, tanto lexical quanto fonética, no sentido de que as mudanças vão afetando as palavras pouco a pouco e pode ocorrer, também, modificação ou perda gradual dos gestos articulatórios (Cf. PAGLIUCA; MOWREY, 1987; BROWMAN; GOLDSTEIN, 1992). Segundo o Modelo de Redes, os vários níveis gramaticais (fonética, morfologia, sintaxe) estão interligados e são modelados de acordo com o uso que se faz da língua.

Ainda com relação a medidas de freqüência, uma questão permanece em aberto: para analisar efeitos de freqüência, é importante definir faixas de freqüência baixa, média e alta para as palavras e tipos (paradigmas) de línguas específicas; no entanto, como os limiares de tais freqüências podem ser determinados? Bybee (2006, p. 06) afirma que não há, até o presente momento, medidas exatas sobre faixas de freqüência para cada um dos fenômenos lingüísticos analisados. No entanto, ela salienta que, conforme os estudos lingüísticos forem evoluindo e aprofundando os recursos estatísticos, faixas mais precisas de freqüência poderão ser determinadas para cada um dos processos lingüísticos pesquisados.

Hay (2001) discute efeitos de freqüência na morfologia e observa que as formas derivadas que são mais freqüentes que suas bases são reconhecidas como palavras inteiras e são semanticamente menos transparentes do que palavras derivadas que são menos freqüentes que suas bases. A autora cita como exemplo a palavra "insane", que tem freqüência de 258 por 17,4 milhões de palavras (Cf. CELEX), enquanto a palavra "sane" tem freqüência de 149 por 17,4 milhões de palavras. "Insane", portanto, é mais freqüente que sua base ("sane") e é, por isso, mais propensa a ser reconhecida e acessada como uma palavra inteira, sendo, também, menos transparente semanticamente, o que significa que suas partes não são decompostas com facilidade. Por outro lado, "infirm" tem freqüência de 27 ocorrências no mesmo corpus, enquanto que "firm" é mais freqüente (715 ocorrências). Então, já que a forma derivada ("infirm") é menos freqüente que sua base ("firm"), a primeira deverá ser mais transparente, facilmente decomposta e acessada através de suas partes (Cf. HAY, 2001, p. 1044). A autora conclui que "because words compete,

the absolute frequency of the derived form is not so important as its frequency relative to the base form with which it is competing." (HAY, 2001, p. 1045). Como essas relações entre freqüência absoluta e relativa são baseadas na maneira como os falantes utilizam a língua, pode-se afirmar que a estrutura do léxico mental e a gramática se modificam de acordo com padrões de uso.

Hay e Baayen (2002) analisam as relações entre freqüência e produtividade. Os autores argumentam que as duas propriedades estão ligadas, e que "Type and token frequency in the lexicon are not good predictors of productivity. But frequency counts of *decomposed* forms in the lexicon *can* predict the degree to which an affix is likely to be productive." (HAY; BAAYEN, 2002, p. 203). Então, não é nem freqüência de tipo nem de ocorrência isoladamente que contam, mas um afixo é produtivo se as palavras que o contêm são decompostas facilmente e o afixo pode ser reconhecido como parte da palavra, separável do radical. Nesse sentido, "The more often words containing a given affix are parsed during perception, the more productive that affix will be." (HAY; BAAYEN, 2002, p. 205). Assumindo-se que a forma como as palavras são acessadas no léxico mental deixa traços na memória e que as palavras estão organizadas em famílias morfológicas, a ativação de um afixo se espalha através de sua rede, fortalecendo a representação mental desse afixo e fazendo com que ele seja mais propenso a ser adotado quando novos itens são flexionados. Então, um afixo se torna mais produtivo à medida que é mais decomposto da palavra durante a produção e reconhecimento desta.

Frisch *et al.* (2001) discutem generalizações fonotáticas no inglês e no árabe. Os autores postulam que, ao contrário do que a teoria gerativa prevê, existe gradiência no julgamento sobre aceitabilidade ou inaceitabilidade de um item léxico dentro dos padrões fonotáticos de uma língua específica. Através de experimentos com o inglês e o árabe, os autores observam que "Very common phonological units contribute positively toward well-formedness, while rare or unattested phonological units contribute negatively." (FRISCH *et al.*, 2001, p. 161); além disso, "Speakers with larger lexicons would accept less probable nonwords, as the probability that an item could occur in a larger lexicon is higher than the probability that an item could occur in a smaller lexicon." (FRISCH *et al.*, 2001, p. 161). Dessa forma, o julgamento dos falantes sobre a (in)existência de palavras no léxico é probabilístico (em vez de categórico), dependendo da freqüência das unidades fonológicas que compõem o item léxico e do tamanho do próprio "dicionário" mental de cada um dos falantes.

Bybee (2001, p. 6-8) estabelece, de forma mais coesa, princípios teóricos importantes relativos ao Modelo de Redes:

1. As representações lingüísticas partilham as mesmas características das representações de outros objetos cognoscíveis, ou seja, informações lingüísticas são armazenadas e acessadas da mesma forma que informações não-lingüísticas, de modo que a mente humana opera da mesma maneira na categorização e processamento de objetos diferentes;
2. A experiência afeta as representações. A forma com que as palavras são usadas modifica sua estocagem e acesso no léxico mental. Nesse sentido, a categorização humana é individual e a forma com que o ser humano experencia todos os eventos que lhe acontecem modela suas representações mentais, não apenas lingüisticamente, mas em termos de conhecimento como um todo;
3. A categorização é baseada em identidade e similaridade, por isso é importante que diferentes ocorrências de uma mesma palavra sejam armazenadas. Com a estocagem de informações redundantes, o léxico mental é capaz de organizar as palavras listadas, considerando, principalmente, a freqüência e a similaridade semântica e fonológica;
4. Não existe separação clara entre léxico e gramática, porque as formas estocadas no léxico já são flexionadas. Generalizações emergem desses itens e novas formas podem ser criadas a partir da analogia, não sendo necessário, portanto, a utilização de regras abstratas para gerar novos itens léxicos.

A partir dos tópicos acima, podemos perceber que o Modelo de Redes defende a importância do uso na dinamicidade dos sistemas lingüísticos. Kemmer e Barlow (2000) discutem algumas características importantes sobre modelos baseados no uso: 1) As unidades lingüísticas não são fixas, mas são remodeladas conforme utilizadas na comunicação diária; 2) Os eventos lingüísticos apresentam a dupla função de manifestarem o conhecimento lingüístico do falante e servirem de material para a reformulação constante do próprio sistema; 3) A freqüência desempenha papel importante na representação mental, visto que ela é um dos grandes responsáveis pela reorganização do sistema lingüístico; 4) Não existe separação formal entre "competência" e "desempenho", porque o desempenho é parte intrínseca e indissociável da competência do falante; 5) O material

primário de estudo lingüístico é a língua usada no cotidiano dos falantes, por isso modelos baseados no uso dependem da análise de diferentes corpora; 6) Uma variação ou mudança lingüística é motivada por três diferentes forças: cognitivas, articulatórias e sociais.

Bybee (1994) discute os conceitos de produtividade, regularidade e fusão. Segundo a autora, em um Modelo de Redes, "productivity is the extent to which a morphological pattern applies to new forms" (BYBEE, 1994, p. 250), "'Regular' is contrasted with 'irregular', where the latter indicates that a pattern is characterized by lexical idiosyncrasies." (BYBEE, 1994, p. 251), e fusão é a falta de transparência entre o radical e o morfema gramatical (BYBEE, 1994, p. 252). Existem algumas relações entre essas propriedades, no seguinte sentido: "morphological or allomorphic processes that are productive tend also to be lexically regular and exhibit a relatively low degree of fusion, while unproductive ones tend to be irregular and characterized by higher fusion" (BYBEE, 1994, p. 252). Apesar disso, essas propriedades não são necessariamente relacionadas, podendo haver casos em que são dissociadas, como ocorre no plural em *–s* do alemão, discutido por Köpcke (1988) e Marcus *et al.* (1995) *apud* Bybee (1994). Em alemão, existem cinco padrões diferentes de pluralização. O paradigma em *–s* tem a freqüência mais baixa (1%) e, conseqüentemente, pouca produtividade (Cf. BYBEE, 1994, p. 259). No entanto, esse padrão em *–s* é usado em metade dos empréstimos lingüísticos mais recentes e em nomes próprios. A preferência por esse morfema de plural ocorre porque ele tem poucas restrições fonológicas e baixo nível de fusão (palavras que se flexionam através dele nunca modificam o radical, por isso a separação entre a desinência de plural e o radical é mais transparente). Então, regularidade e produtividade são independentes neste caso, porque o sufixo menos produtivo (*–s*) é utilizado para flexionar novos itens que entram na língua. Segundo Bybee (1994), não há distinção categórica entre regular/irregular ou produtivo/improdutivo. Esses conceitos são pontos extremos de um mesmo *continuum* e não são essas propriedades inerentes à palavra que determinam se ela será ou não listada no léxico. Existem duas formas de acesso a uma palavra: lexicalmente, se ela tiver alta freqüência de ocorrência, e através de sua classe, se sua freqüência de ocorrência for baixa. Portanto, o que define a estocagem de um item é sua freqüência de ocorrência, não sua estrutura interna (regular ou irregular). Por conta disso, regularidade não significa, necessariamente, produtividade mais alta. Observamos, então, que o Modelo de Redes propõe categorizações baseadas em medidas de freqüência determinadas pelo uso da língua.

Bybee e Slobin (1982) observam que verbos irregulares do inglês, apesar de suas idiossincrasias, são altamente freqüentes na língua, o que garante sua autonomia lexical e faz com que mantenham suas peculiaridades. Quanto a esses verbos (irregulares), os autores afirmam que a memorização é imprescindível, visto que relações como "go"/"went" não poderiam ser aprendidas por regras. No entanto, mesmo entre itens irregulares, algumas generalizações podem ser feitas, alocando as palavras em classes menores em função das similaridades existentes entre as formas flexionadas. Assim, no inglês, existem, por exemplo: 1) Verbos que mantêm a mesma forma no presente e no passado, como "beat" e "cut"; 2) Verbos que mudam /d/ final para /t/, como "sen<u>d</u>"/"sen<u>t</u>" e "buil<u>d</u>"/"buil<u>t</u>". Dessa forma, os verbos irregulares são estocados no léxico e as similaridades que eles partilham fazem com que se organizem em esquemas com redes de interconexões semânticas e fonológicas.

Bybee e Moder (1983) analisam subclasses de verbos irregulares do inglês ("swim"/"swam"/"swum" e "string"/"strung") e tentam responder a duas questões: 1) Que critérios os falantes usam para classificar determinadas palavras como pertencentes a uma mesma classe? e 2) Por que algumas classes de palavras atraem novos membros e outras desaparecem?

Com relação à definição de categorias, Bybee e Moder (1983, p. 256) afirmam:

> Wittgenstein 1953 proposed that the categories which we create construing our universe do not consist of sets of items sharing a common set of attributes not shared by other items, but rather that members of categories form chains of family resemblances. […] Rosch & Mervis 1975 and Rosch 1978 develop this idea further by proposing that each category has a PROTOTYPE or best exemplar, which has a cluster of attributes, and more marginal members, with perhaps only one attribute that puts them in the category.

Portanto, a categorização de palavras em uma mesma classe é feita a partir de similaridades com um membro prototípico, o que não significa que todos os exemplares devam partilhar, obrigatoriamente, as mesmas características.

Com relação ao fato de algumas classes de palavras atraírem novos membros e outras desaparecerem da língua, Bybee e Moder (1983) defendem que isso é conseqüência da produtividade, ou seja, da capacidade que uma classe tem para atrair novos membros ao longo dos séculos. Para que uma classe possa abarcar novos itens, é necessário que ela não tenha muitas restrições de ordem fonológica e possa, portanto, se estender a novas palavras.

Bybee e Newman (1995) discutem dois processos morfológicos diferentes: afixação (adoção de uma desinência, como passado, ou de sufixos e prefixos) e mudança de radical (flexões que ocorrem não a partir de afixos, mas com modificações no radical da palavra). É comumente afirmado que o primeiro processo (afixação) é mais natural[49] que o segundo (mudança de radical). Bybee e Newman (1995) elaboraram um experimento para testar essa afirmação. Os autores criaram uma língua artificial, em que metade dos nomes era flexionada através de mudança de radical e a outra metade, através de sufixação. Depois de uma fase de aprendizagem, em que palavras dessa língua foram ensinadas a voluntários, os falantes foram convidados a formar o plural de palavras novas, tendo como base o que eles haviam aprendido sobre a língua artificial.

Nos resultados do experimento, os falantes não demonstraram nenhuma preferência por sufixação em detrimento de mudança de radical. Houve 48,75% de respostas com sufixos e 50,25% de respostas com mudança de radical, ou seja, as porcentagens para os dois casos foram muito próximas. Já que as palavras ensinadas aos falantes na fase de aprendizagem eram metade flexionada por sufixação e metade por mudança de radical, parece que essa igualdade de freqüência de tipo teve um efeito significativo nas respostas dos falantes. Os resultados evidenciam que o uso da língua desempenha um papel importante na representação mental do conhecimento lingüístico e, mais que isso, na maneira como padrões existentes são aplicados a novos itens. A conclusão a que Bybee e Newman (1995) chegam é que palavras inteiras são estocadas no léxico e, além disso, ambos os processos (sufixação e mudança de radical) são igualmente passíveis de aplicação a novos itens, não havendo vantagem de um sobre o outro. O que faz com que uma estrutura seja estendida a novas palavras não é o processo em si mesmo, mas sua freqüência de tipo. Se essa for alta, o paradigma será usado em novos itens; se não for, o padrão será restrito apenas à sua classe originária.

Bybee (2006, p. 02) afirma que, em uma teoria baseada no uso, a gramática pode ser definida como "[...] the cognitive organization of one's experience with language." Nessa perspectiva cognitivista, a gramática não é vista como um sistema estático, mas como uma estrutura que emerge, sobretudo, da repetição de eventos comunicativos na comunicação diária. Assim, a freqüência com que estruturas ou palavras ocorrem tem o

---

[49] A suposta "naturalidade" de sufixação com relação a mudança de radical se deve aos seguintes fatores: 1) Afixação é mais comum nas línguas naturais e mais recorrente dentro de línguas específicas; 2) Na aquisição lingüística, afixação é aprendida antes que mudança de radical; 3) Afixação é mais fácil de processar do que mudança de radical (Cf. BYBEE; NEWMAN, 1995).

papel de remodelar a gramática, fazendo com que novas estruturas emerjam e antigas caiam em desuso.

A noção de freqüência e seus possíveis efeitos nos fenômenos lingüísticos ganhou destaque com o Modelo de Redes. No entanto, na literatura lingüística, efeitos de freqüência já haviam sido analisados em alguns trabalhos anteriores. A título de esclarecimento, mencionaremos, brevemente, as abordagens de Schuchardt (1885), Leslau (1969), Fidelholtz (1975) e Phillips (1984).

Schuchardt (1885, p. 58), comentando vários aspectos relacionados à mudança lingüística, faz a seguinte afirmação: "Rarely-used words drag behind; very frequently used ones hurry ahead. Exceptions to the sound laws are formed in both groups." Esse comentário é bastante importante, visto que, há mais de um século, o autor já noticiava que palavras pertencentes a diferentes faixas de freqüência seriam afetadas de forma distinta pela mudança lingüística.

Leslau (1969), estudando as línguas etíopes, analisa mudanças fonéticas que afetam as palavras mais freqüentes primeiro e as menos freqüentes depois. Em sua pesquisa, o autor observa que, nas línguas investigadas, houve várias mudanças fonéticas, como aspiração, elisão, assimilação, etc., e todas elas afetaram as palavras mais freqüentes dentro daquela realidade cultural: "comer", "Deus", "dizer", "fazer", "igreja", "poder" (verbo), "saber", "vestir".

Fidelholtz (1975) aponta o fator freqüência de ocorrência como determinante da suscetibilidade à mudança sonora. Analisando a redução vocálica no inglês, o autor observa uma correlação entre familiaridade da palavra e estocagem lexical. Uma palavra familiar é muito usada, o que a tornará mais acessível no léxico mental e, conseqüentemente, fará com que seja mais suscetível a mudanças sonoras. Segundo o autor, as palavras freqüentes e as infreqüentes são estocadas em áreas distintas do cérebro, por isso exibem diferente propensão à mudança.

Phillips (1984) postula que a mudança sonora afeta alguns itens lexicais antes de outros e, além disso, a mudança não opera somente com base no condicionamento fonético. Para Phillips (1984), as mudanças que atingem as palavras mais freqüentes primeiro incluem redução vocálica, apagamento ou assimilação, ou seja, mudanças baseadas na fisiologia da fala. A autora ressalva que, com essa afirmação, não está pressupondo que somente os fatores fonéticos geram a mudança, mas salienta que a influência da superfície fonética é imprescindível para tais mudanças. Posteriormente, com relação às mudanças

que afetam as palavras menos freqüentes primeiro, a autora postula que estas são mudanças não-fisiologicamente motivadas. Então, a hipótese de Phillips (1984) para efeitos de freqüência na estrutura lingüística pode ser sumarizada da seguinte forma: mudanças fisiologicamente motivadas afetam as palavras mais freqüentes primeiro, enquanto que mudanças não-fisiologicamente motivadas afetam as palavras menos freqüentes primeiro.

Com base nos comentários dos autores supracitados, percebemos, portanto, que a noção de freqüência do item lexical já era considerada na literatura lingüística antes que o Modelo de Redes fosse elaborado. As pesquisas de Bybee definiram de forma mais sistemática como a freqüência (de tipo ou de ocorrência) interfere nas representações mentais e tem o papel de remodelar a estrutura lingüística; no entanto, investigações anteriores já noticiavam a importância da freqüência nos processos de variação e mudança lingüística.

Comparando os pontos comuns e discordantes entre o Modelo de Redes e os modelos resenhados anteriormente (Modelo Conexionista e Teoria de Palavras e Regras), Bybee (1995) afirma que o modelo proposto por ela (Modelo de Redes) partilha semelhanças com o Modelo Conexionista, uma vez que ambos conferem à freqüência de tipo um papel importante na produtividade de um padrão lingüístico. Por outro lado, ambos os modelos (Conexionista e de Redes) diferem da Teoria de Palavras e Regras, visto que esta última restringe aos itens irregulares o papel da freqüência de ocorrência.

O Modelo de Redes também se diferencia da Teoria de Palavras e Regras em outros pontos mais específicos. Em primeiro lugar, conforme vimos, a Teoria de Palavras e Regras pressupõe que a armazenagem de informações lingüísticas é tão peculiar que possui um módulo específico no cérebro, por isso a linguagem é processada de maneira diferente dos demais objetos cognoscíveis. Ao contrário disso, o Modelo de Redes propõe que a mente humana é um todo organizado, que possui uma única forma de lidar com as informações advindas do meio externo, sejam elas lingüísticas ou não. Um segundo ponto de discordância entre essas teorias é o fato de que, para o Modelo de Redes, o uso da língua em situações concretas de comunicação afeta as representações mentais de informações lingüísticas. Teorias como a de Pinker (1991) desconsideram a importância dos fatores pragmáticos e sociais e tratam a gramática como algo estático, não suscetível a variações por conta do uso cotidiano da língua.

Mencionando as similaridades entre os diferentes modelos de léxico mental, Bybee (1995, p. 432) afirma que, no Modelo de Redes e no Modelo Conexionista, flexões

regulares e irregulares são processadas da mesma forma e não existem regras morfológicas explícitas. O fato de que diferentes classes morfológicas apresentam níveis distintos de produtividade deriva não de dois mecanismos diferentes de processamento de itens léxicos (palavras e regras), mas é fruto da interação entre freqüência de tipo e características que fazem com que um determinado esquema seja mais ou menos receptivo a novas palavras, conforme discutimos anteriormente. Como a unidade de armazenamento lexical é a palavra inteira, primitiva ou derivada, no Modelo de Redes não existe, em termos formais, separação entre léxico e gramática.

O Modelo Conexionista e o Modelo de Redes apresentam similaridades por postularem que apenas um mecanismo associativo é responsável por flexões morfológicas, sejam elas regulares ou irregulares. Apesar dessas semelhanças, existem, também, diferenças entre esses dois modelos:

> […] the treatment of token frequency in connectionist models is taken to be the frequency of the mapping between base and derived form […]. In the network model, it is the frequency of the derived word itself that is significant and is represented as lexical strength, not a mapping between two forms. In fact, the higher the frequency of the derived form, the weaker the mapping between it and the basic form. (BYBEE, 1995, p. 432)

Portanto, no Modelo Conexionista, a freqüência de ocorrência é considerada em função do mapeamento existente entre a forma base e a derivada; no Modelo de Redes, por sua vez, a freqüência de ocorrência é a freqüência da forma em si mesma, seja ela primitiva ou derivada. Sendo assim, no Modelo de Redes, quanto maior a freqüência de ocorrência de um item lexical, mais fracas são as conexões entre ele e a rede morfológica de que faz parte.

Outra divergência entre o Modelo de Redes e o Conexionista é o fato de que as redes neurais propõem o mapeamento apenas entre a forma primitiva e a derivada. O Modelo de Redes, por sua vez, estabelece a possibilidade de generalizações a partir de formas derivadas. Existiria, portanto, mapeamento entre formas primitivas e derivadas, mas haveria, também, mapeamento de formas derivadas entre si. O Modelo Conexionista não prevê esse tipo de possibilidade de generalização. Esses seriam, portanto, pontos de discordância entre o Modelo de Redes e o Conexionismo.

De acordo com o que vimos ao longo de toda esta subseção, o Modelo de Redes propõe a incorporação do uso da língua como "remodelador" das representações mentais e da estrutura lingüística. Essa perspectiva plástica de léxico mental nos leva à

conceituação de gramática como algo nunca completo, mas sempre em processo de construção, de acordo com a cultura com a qual a língua interage:

> The grammar of a language, then, consists not of a single delimited system, but rather, of an open-ended collection of forms that are constantly being restructured and resemanticized during actual use. […] Our speech is a vast collection of hand-me-downs that reaches back in time to the beginning of language. The aggregation of changes and adjustments that are made to this inheritance on each individual occasion of use results in a constant erosion and replacement of the sediment of usage that is called grammar. (HOPPER, 1998, p. 159)

Conforme esclarecemos anteriormente, uma das opções teóricas adotadas por nossa pesquisa é o Modelo de Redes. Em linhas gerais, essa teoria proporia as seguintes hipóteses para os plurais sob análise nesta tese:

- Como é a freqüência de tipo que determina a produtividade de uma classe, pode-se supor que: i) Para os itens terminados em *–ão* no singular, haverá um predomínio da flexão através do morfema *–ões*, que é o mais freqüente para a classe de palavras terminadas em *–ão* no singular; ii) Para os itens terminados em *–l* e *–u* no singular, o padrão de pluralização em *–is* pode ser adotado para pluralizar palavras terminadas em ditongo em *–u*, já que, perdida a distinção fonética entre *–l* e *–u* em fim de sílaba, o tipo em *–l* é mais freqüente que o tipo em *–u* para pluralizar itens terminados em [w];
- Já que não existe prevalência de afixação sobre mudança de radical, os falantes não devem demonstrar nenhuma tendência favorecendo a adoção do plural em *–ãos* (aplicação do morfema *–s*) em detrimento de *–ães* ou *–ões* (mudança de radical). De forma análoga, não deve haver, também, favorecimento do plural em *–us* (regular) sobre a forma em *–is* (com alteração no radical), para itens terminados em *–l* ou *–u*, já que não há distinção fonética entre esses segmentos e, por isso, em teoria, a pluralização em *–us* ou *–is* pode ser aplicada;
- Como a freqüência de ocorrência é responsável pela manutenção de irregularidades, é possível que palavras com alta freqüência de ocorrência preservem seus plurais etimológicos, já que sua repetição contribuiria para sua força lexical e garantiria sua permanência no léxico;

- Já que as palavras, no léxico mental, se organizam a partir de similaridades fonológicas e semânticas, supõe-se que itens monossílabos, como "mão", "pão", "sal", "sol", "réu", "véu", sejam mais resistentes à variação, por formarem uma classe mais coesa e definida em termos fonológicos;
- O modelo não postula, necessariamente, a atuação de regras para gerar formas de superfície, então é possível que generalizações emerjam a partir de classes específicas de plural.

Em linhas gerais, a partir das hipóteses levantadas pelo Modelo de Redes, os plurais sob análise nesta tese poderiam ser explicados como um reordenamento das representações mentais a partir da interação entre as freqüências de tipo e de ocorrência, decorrente do uso da língua em situações de comunicação cotidiana.

A outra proposta teórica a ser adotada por esta tese é o Modelo de Exemplares, discutido a seguir.

4.4 Modelo de Exemplares

O Modelo de Exemplares[50] ("Exemplar Model" ou "Exemplar Theory") foi primeiramente desenvolvido como um modelo de similaridade e classificação no campo da psicologia. Posteriormente, ele foi aplicado à lingüística, na área de reconhecimento de fala, primeiramente por Johnson (1997) e Lacerda (In press).

Basicamente, o Modelo de Exemplares propõe inovações que decorrem de críticas a padrões gerativos de fonética e fonologia. Segundo Pierrehumbert (2000, 2001a, 2001b, 2001c, 2003), em modelos gerativos:

- O léxico é separado da gramática fonológica (Cf. CHOMSKY; HALLE, 1968; CHOMSKY; LASNIK, 1995 *apud* PIERREHUMBERT, 2001b) e a produção lingüística se dá da seguinte forma: uma palavra é recuperada no léxico e regras da

[50] Mais uma vez, para garantir a coerência da nomenclatura em português, vamos adotar os termos "Modelo de Exemplares" ou "Teoria dos Exemplares".

gramática a transformam na forma de superfície que ela deve apresentar. As regras são abstratas, então se aplicam não a itens léxicos isoladamente, mas a todas as palavras que partilham determinadas características. Não existe, portanto, espaço para que algumas palavras se submetam à regra e outras não;

- É desconsiderada a noção de freqüência. De acordo com tais teorias, freqüência faz parte do desempenho lingüístico, não da competência. A questão da alofonia, por exemplo, é considerada em termos de distribuição complementar, não de freqüência de ocorrência dos itens lexicais;
- O conhecimento implícito do falante sobre os sons da língua resume-se ao conhecimento dos sons propriamente ditos e das regras abstratas para combiná-los e garantir a boa-formação das palavras. Nesse sentido, os julgamentos dos falantes sobre a existência de determinada palavra (ou seqüência de sons) na língua são categóricos: o item é existente ou inexistente na estrutura lingüística. Não existe, portanto, a possibilidade de julgamentos graduais entre esses dois extremos;
- Variações fonéticas relacionadas a registros sociais e estilísticos não são incluídas na análise lingüística: "Though the authors of such models would no doubt acknowledge the existence of such variation, they have not undertaken to provide a formal treatment of the cognitive capabilities which permit it." (PIERREHUMBERT, 2001a, p. 1)

Segundo Pierrehumbert (2000, 2001a, 2001b, 2001c, 2003), alguns fenômenos lingüísticos desafiam o escopo teórico de tais modelos:

- Não existe um simples caso em que fonemas análogos em línguas diferentes possuam exatamente a mesma realização fonética. Até mesmo palavras diferentes dentro da mesma língua apresentam realizações fonéticas distintas para fonemas iguais. Hooper (1976), analisando a redução de schwa, observou que, em palavras de alta freqüência, como "every" e "evening", o schwa era completamente ausente, juntamente com a sílaba em que este se encontrava; em palavras de média freqüência, como "memory" e "salary", o schwa era apagado, mas o /r/ permanecia; em palavras de baixa freqüência, como "mammary" e "artillery", o schwa se mantinha e a sílaba permanecia inalterada. Tais dados sugerem que o detalhamento fonético é relacionado a palavras específicas. Isso desafia as regras

propostas pela teoria gerativa, pois essas deveriam ser aplicadas a todas as palavras que apresentam o mesmo contexto fonético. A armazenagem de detalhes fonéticos em termos graduais não é prevista na teoria gerativa;

- Ainda tomando como exemplo o fenômeno analisado por Hooper (1976), observamos que houve uma gradualidade entre apagamento e presença do schwa com relação à freqüência de ocorrência dos itens lexicais, no sentido de que palavras freqüentes sofreram taxas mais altas de apagamento e palavras infreqüentes preservaram mais o schwa. Assim, percebe-se que a freqüência dos itens lexicais tem efeito não só no desempenho lingüístico, mas também na competência. Se palavras que partilham contextos fonéticos semelhantes ("memory" e "mammary", por exemplo) não apresentam as mesmas taxas de redução de schwa, é sinal de que essa distinção, ocasionada pela freqüência de ocorrência de tais palavras, está listada no léxico mental, em vez de ser derivada a partir de regras abstratas para gerar a forma de superfície;
- Com relação às condições de boa-formação da sílaba na língua (fonotática), Treiman *et al.* (2000 *apud* PIERREHUMBERT, 2001c) observaram que, em uma tarefa de julgar a aceitabilidade ou não de palavras inventadas, os falantes apresentaram o seguinte resultado: palavras que continham combinações de sílabas que, na sua língua nativa, eram mais freqüentes foram consideradas melhor formadas do que palavras que continham combinações de sílabas que, na língua dos falantes, eram infreqüentes. Esse resultado mostra que, em vez de haver distinção categórica entre palavra possível e impossível de ocorrer em uma língua, os indivíduos procedem por um julgamento gradiente e probabilístico, que toma como critério as estatísticas do léxico. Assim, um modelo de léxico mental deve admitir a possibilidade de categorias intermediárias entre possível/impossível de ocorrer na língua;
- Análises sociolingüísticas (LABOV, 1972, 1994, 2001) comprovam que fatores sociais e estilísticos, tais como faixa etária, classe social, nível de escolaridade e de formalidade da interação verbal, interferem na adoção de diferentes padrões de fala.

A partir dos pontos mencionados acima, parece claro que uma abordagem categórica, baseada em regras simbólicas e que exclui os fatores sociais e pragmáticos das representações mentais não consegue justificar a diversidade de fenômenos encontrada nas

línguas do mundo. Tomando como ponto de partida as críticas à gramática gerativa, o Modelo de Exemplares apresenta uma proposta alternativa de léxico mental e propõe que se considere (Cf. PIERREHUMBERT 2000, 2001a, 2001b, 2001c, 2002):

1) A armazenagem de detalhamento fonético no léxico mental, já que, conforme mencionamos anteriormente, os mesmos fonemas apresentam realizações fonéticas distintas em línguas diferentes ou até dentro de uma mesma língua;
2) A noção de freqüência, porque ela apresenta efeitos na memória de longo termo e é a partir dela que categorias são organizadas no léxico mental;
3) A noção de gradualidade nas representações mentais, no sentido de que as distinções entre os segmentos não são categóricas, em termos de presença/ausência de um traço, mas tais diferenças são distribuídas em um *continuum* no espaço articulatório e acústico;
4) Fatores sociais, já que esses interferem nas representações mentais e os falantes categorizam, também, informações não-lingüísticas sobre os itens léxicos.

O Modelo de Exemplares postula que, na aquisição do sistema fonético, o falante categoriza em termos probabilísticos o sinal de fala recebido; sendo assim, padrões mais freqüentes são mais acessíveis e, portanto, aprendidos primeiro, ao passo que categorias menos usadas ou com poucos exemplares são construídas mais tarde. Para o Modelo de Exemplares, a aquisição do inventário fonético da língua ocorre através da generalização estatística de padrões disponíveis no ambiente no qual a criança convive. Devido à variabilidade intrínseca à fala, é imprescindível que detalhes fonéticos sejam adquiridos e que generalizações estatísticas organizem tais informações em termos de membros centrais e periféricos no léxico mental.

Quando uma criança está adquirindo a língua, se depara com uma grande quantidade de dados variáveis, então, em vez de abstrair regras simbólicas a partir desses dados (conforme proposto pela teoria gerativa), desenvolve um sistema cognitivo de representações em que a freqüência com que certas estruturas lingüísticas ocorrem desempenha papel importante. A noção de gramática não é, entretanto, perdida:

> The cognitive system is still grammatical: It establishes the well-formedness of complex forms from their subparts, and it has the power to create and to process completely novel forms. However, it is a probabilistic grammar, in the sense that it maintains frequency distributions, and the frequency of any given phonological unit is an important factor in how it behaves in the system. (PIERREHUMBERT, 2001c, p. 195-6)

Ainda com relação à aprendizagem da língua, Pierrehumbert (2001c, p. 196) afirma que a aquisição tardia de alternâncias morfofonológicas deriva do fato de que estas devem ser deduzidas a partir da comparação de pares de palavras. A ligação entre tais itens não ocorre até que o falante tenha experienciado bastante a ocorrência das duas palavras (a primitiva e a derivada) e estabelecido as devidas relações entre elas. Por causa disso, algumas dessas alternâncias são adquiridas mais tarde.

No Modelo de Exemplares, a organização lexical se dá da seguinte maneira:

> [...] each category is represented in memory by a large cloud of remembered tokens of that category. These memories are organized in a cognitive map, so that memories of highly similar instances are close to each other and memories of dissimilar instances are far apart. (PIERREHUMBERT, 2001b, p. 140)

Se cada ocorrência nova de uma palavra é armazenada na memória, pode-se argumentar que a mente humana não teria capacidade de estocagem suficientemente grande para guardar todas as palavras ouvidas ao longo de uma vida inteira. O Modelo de Exemplares responde a essa crítica da seguinte maneira: 1) A mente humana apresenta grande capacidade de memória, o que trabalhos como o de Johnson (1997) comprovam; 2) A memória decai gradualmente, de modo que palavras ouvidas há alguns dias são relembradas mais facilmente do que itens ouvidos há muitos anos; 3) Mesmo com a categorização de vários exemplares e de seu detalhamento fonético, "the parameter space in which the exemplars are represented is assumed to be granularized. Exemplars whose differences are too fine to show up under the granularization are encoded as identical [...]" (PIERREHUMBERT, 2001b, p. 141). Assim, um exemplar individual não corresponde a uma experiência única, mas a uma classe de experiências que foram percebidas como praticamente idênticas. Portanto, uma "nuvem de exemplares" é formada a partir de vários exemplares similares da mesma categoria.

Com relação à classificação e estocagem lexical, o Modelo de Exemplares prevê que, quando o falante encontra uma nova ocorrência de um item, este é classificado de acordo com as similaridades que partilha com outros itens já estocados. A codificação do

novo item depende das relações já existentes nas redes de exemplares. Em geral, um grupo que tem mais exemplares vence a competição pela categorização do novo item, porque a alta freqüência de tipo reforça a nuvem de exemplares e acaba por atrair mais membros. Como os exemplares estão organizados em nuvens, o acesso a um exemplar ativa, conseqüentemente, todos os itens com que este mantém alguma relação de similaridade. A categorização e a armazenagem dos itens no léxico mental depende de uma grande quantidade de experiência e exposição a uma determinada língua. A organização das representações lingüísticas é uma tarefa que começa na infância, atinge seu pico na adolescência, mas dura toda a vida, porque a incorporação de novos exemplares através de experiências vividas faz parte da cognição humana.

Pierrehumbert (2001a, p. 07) comenta o seguinte sobre o Modelo de Exemplares:

> The critical ingredients in exemplar theory are a map of the perceptual space and a set of labels over this map. […] For phonetics, the relevant physical domain is the articulatory/acoustic space, whose dimensions are the relevant dimensions of contrast in articulation and acoustics.

Na produção de um exemplar, um alvo "is obtained by picking an exemplar randomly from the exemplar list of the desired label. Since the probability (or strength) of each exemplar is time-dependent, old exemplars are only rarely used as targets." (PIERREHUMBERT, 2001a, p. 09). Dessa forma, a freqüência desempenha papel crucial, visto que os exemplares mais recorrentes apresentam representações mentais mais fortes e, conseqüentemente, têm maior probabilidade de serem utilizados.

Pierrehumbert (2003, p. 182) menciona a pesquisa realizada por Peterson e Barney (1952). Os autores coletaram dados de vogais pronunciadas por diferentes falantes do inglês e, a partir dessas informações, elaboraram a seguinte figura, mostrando a distribuição das vogais no espaço acústico/articulatório:

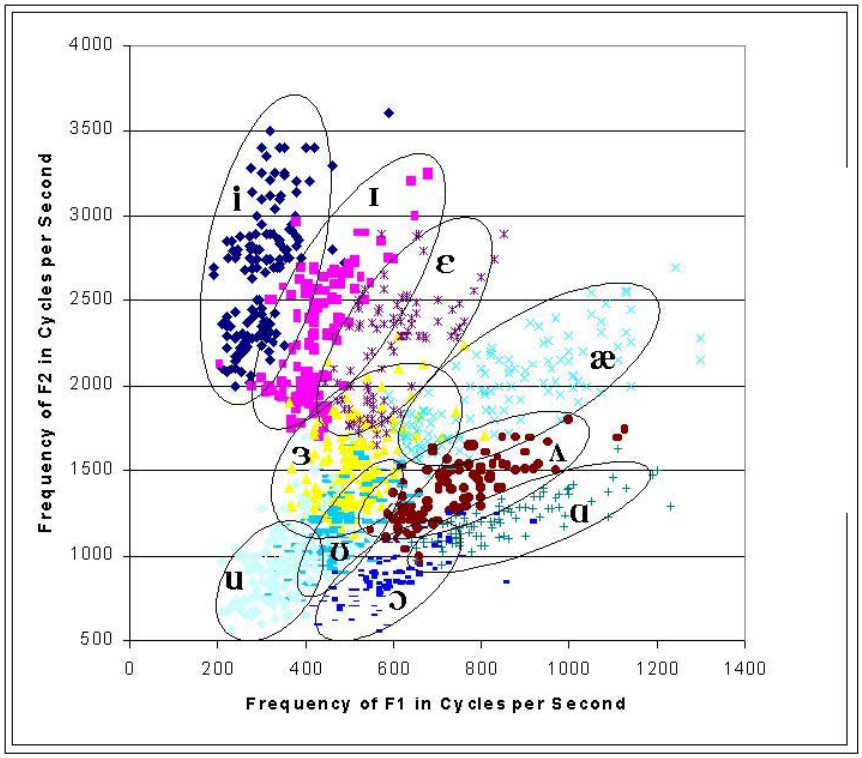


Figura 4: O espaço acústico/articulatório das vogais do inglês

Fonte: Pierrehumbert, 2003, p. 183.

Nesta figura, podemos perceber que as vogais apresentam distribuições probabilísticas ao longo de F1 e F2. Em vez de haver pontos específicos de localização para as vogais, cada uma delas ocupa uma região contínua no espaço fonético. Observamos, também, que as vogais concentram-se mais no centro das categorias do que nas bordas. Além disso, como a distribuição das vogais é probabilística e não absoluta, existem casos de sobreposição de segmentos, em que exemplares de diferentes vogais estão alocados em um mesmo espaço articulatório/acústico. Essa classificação das vogais nos leva à conclusão de que a realização fonética de um segmento é sempre variável. Ainda que um mesmo falante pronuncie o mesmo fonema em um mesmo contexto, haverá variação, inevitavelmente. Essa variabilidade inerente à fala é tão grande que os símbolos fonéticos propostos pelo IPA (International Phonetic Alphabet) não são capazes de representar perfeitamente todas as diferentes possibilidades de pronúncia para o mesmo segmento. Nesse sentido, o IPA deve ser entendido apenas como uma abstração dos segmentos

possíveis nas línguas do mundo, de acordo com seus traços articulatórios, ou como o centro (exemplar prototípico) de uma categoria variável.

No mapeamento da nuvem de exemplares, fatores como diferenças dialetais, peculiaridades de cada falante e variação livre da fala também são consideradas. Bybee (2001, p. 52), citando Johnson (1997), apresenta uma nuvem de exemplares com a inclusão de fatores lingüísticos e extralingüísticos:

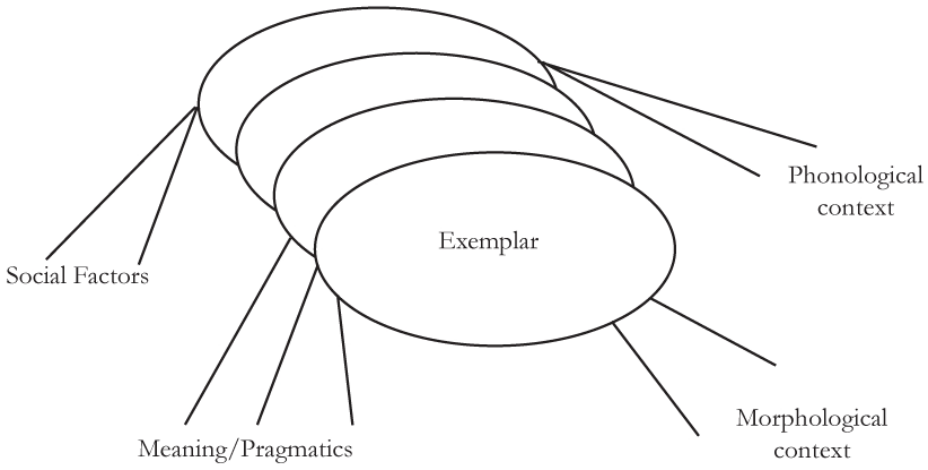


Figura 5: Nuvem de exemplares

Conforme o esquema acima demonstra, a nuvem de exemplares comporta informações sociais e lingüísticas, visto que é no contexto de interação social que os padrões lingüísticos constituintes da nuvem de exemplares são adquiridos.

Já que existe uma nuvem de exemplares com todas as variações já experenciadas para cada categoria, a escolha do exemplar a ser utilizado na fala é uma tarefa que depende, principalmente, da freqüência das formas no léxico mental. Dessa forma, itens que são ouvidos constantemente tendem a ser produzidos com freqüência, uma vez que suas representações são mais fortes na mente do falante. No entanto, já que fatores sociolingüísticos são categorizados, pode acontecer, também, que determinados exemplares, embora armazenados, sejam evitados por um falante, porque algumas formas carregam juízos negativos de valor.

Com relação aos plurais do PB, poderíamos estabelecer uma nuvem de formas em competição:

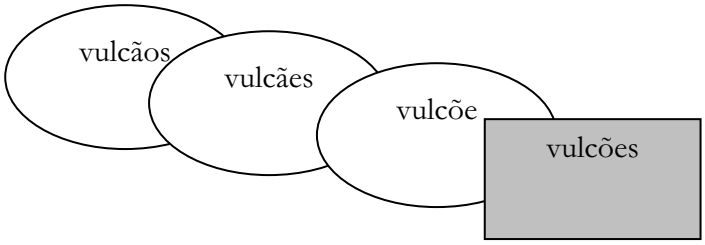


Figura 6: Itens em competição na nuvem de exemplares

Na nuvem de exemplares acima, podemos observar a representação mental das possibilidades de plural para a palavra "vulcão". Naturalmente que esses exemplares só se aplicam no caso de a pessoa já ter ouvido todas essas formas. Devemos notar que até mesmo formas não-padrão, como "vulcõe", estão estocadas, porque, uma vez que o falante ouviu essa variação, ela faz parte de sua experiência e é armazenada no léxico mental. O fato de a palavra "vulcões" ser apresentada em um formato diferente e em sombreado tem o objetivo de enfatizar que esse é o exemplar mais forte da categoria. Como "vulcões" é o mais freqüente na fala, tende a ser, também, o mais recorrente na produção, visto que sua representação mental é mais proeminente. Isso não impede que as demais formas sejam utilizadas esporadicamente, já que estão disponíveis no léxico mental. No entanto, outra conseqüência da competição dessas formas pode ser o fato de "vulcões" se tornar freqüente a ponto de excluir do léxico mental as demais formas. Nesse caso, teríamos a passagem de uma mudança em progresso para uma mudança concluída (pelo menos até certo estágio, o que não impede que outras variações atinjam uma forma recém--categorizada após mudança em progresso[51]). As representações mentais estão sempre sendo atualizadas no dia-a-dia do falante. A cada conversação, novos itens são acrescidos às nossas nuvens de exemplares. Conforme já mencionamos anteriormente, essa tarefa de categorização dos itens léxicos dura a vida inteira do falante.

Outro ponto importante assumido pelo Modelo de Exemplares é a questão da variabilidade na configuração dos segmentos em línguas diferentes. Pesquisas recentes demonstram que o mesmo segmento – por exemplo, /p/ – tem propriedades distintas em duas línguas diferentes e pode variar, também, dentro de uma mesma língua. Tais características podem mudar em detalhes como nível de vozeamento, aspiração ou até em

[51] "Mudança em progresso" seria a adoção gradual de uma estrutura lingüística nova em detrimento de uma antiga. Em geral, esse tipo de mudança pode ser percebido através da diferenciação entre a fala de pessoas jovens e idosas, no sentido de que os jovens tendem a utilizar a forma inovadora com mais freqüência e os idosos tendem a preservar a antiga. "Mudança concluída" é um estágio posterior ao da mudança em progresso, quando não há mais duas formas lingüísticas competindo porque uma delas já se sobrepôs à outra.

traços supra-segmentais. Isso sugere que o falante aprende detalhamento fonético de palavras específicas como parte intrínseca do inventário fonético de sua língua. Sendo assim, o Modelo de Exemplares nega a abordagem tradicional, que postula a separação entre léxico e gramática e defende que apenas traços não-redundantes são estocados no léxico:

> [...] the model assumes that detailed phonetic memories are associated with individual words and it implicitly defines word specific probability distributions over phonetic outcomes. Whereas the classic models define a strong separation between the lexicon and the grammar, in the present model these represent two degrees of generalization over the same memories and are thus strongly related to each other. (PIERREHUMBERT, 2001b, p. 139)

A cadeia sonora contínua da fala faz com que os gestos articulatórios dos segmentos de uma palavra entrem em contato uns com os outros, variando os traços peculiares aos fonemas. Isso parece indicar que o armazenamento de informações lingüísticas comporta detalhamento fonético (Cf. JOHNSON; MULLENIX, 1997). Por causa disso, a noção de alofonia ganha outras dimensões, já que não existe um fonema com características fixas *a priori* e um alofone em distribuição complementar. Na verdade, a variação é presente em todos os segmentos, dependendo de quem fala, em que registro social, bem como do contexto fonético dos segmentos pronunciados. É interessante notar, também, que, a despeito de toda essa variabilidade inerente à sua língua materna, as crianças, de alguma forma, conseguem normalizar essa fala, categorizando segmentos que partilham peculiaridades articulatórias e acústicas.

Comentando a questão do armazenamento de detalhes fonéticos, Goldinger (1996) analisa dados sobre a percepção de vozes distintas em diferentes intervalos de tempo. O autor observa que, em um experimento em que palavras eram repetidas pela mesma voz ou por vozes diferentes, os falantes identificaram melhor as palavras faladas pela mesma pessoa. Isso parece ser um indício de que o léxico mental é capaz de armazenar detalhes sobre a voz do sujeito falante.

Johnson (1997) enfatiza que um falante pode determinar o gênero de alguém ao ouvir uma voz, assim como é, também, capaz de reconhecer vozes de pessoas conhecidas, sem visualizar suas feições. Tais fatos sugerem que as propriedades acústicas que diferenciam pessoas são guardadas na memória do falante, por isso o autor postula a necessidade de um Modelo de Exemplares que acomode a representação de detalhamento fonético na mente do indivíduo. Em tal modelo, as categorizações ocorrem em termos de

freqüência de ocorrência dos itens lexicais, de modo que palavras mais freqüentes apresentam níveis de ativação maiores que palavras infreqüentes. No modelo proposto pelo autor, já que o léxico mental comporta detalhamento fonético, a fala é percebida e categorizada sem a necessidade de regras abstratas que realizem sua normalização.

Hume e Johnson (2001) discutem a interface entre percepção da fala e fonologia. Analisando questões relacionadas à variação e mudança lingüística, os autores afirmam que existem diferentes possibilidades de mudanças a que os itens lingüísticos estão suscetíveis, tais como epêntese, dissimilação, metátese, assimilação e apagamento. Dentre essas (e outras) possibilidades de mudança, o que determina o caminho que uma forma lingüística adotará é um "filtro" composto pelos seguintes fatores: **percepção** (audição e reconhecimento), **produção** (coordenação e aerodinâmica), **generalização** (categorias cognitivas que definem a representação do sistema fonético na mente dos falantes) e **conformidade** (comunicação e sociedade). A interação entre tais fatores determina o caminho de uma forma lingüística *p* a uma forma lingüística *p'* (HUME; JOHNSON, 2001, p. 10). Os autores enfatizam, também, o fato de que os mesmos princípios que regulam a mudança lingüística através dos anos atuam na comunicação diária, remodelando a linguagem no cotidiano.

Pierrehumbert (2001b, p. 143) menciona dois sucessos alcançados pelo Modelo de Exemplares: i) Já que, para esse modelo, o detalhamento fonético sobre palavras individuais é parte indissociável da aquisição lingüística e da categorização de novos itens, o modelo fornece explicações plausíveis para o fato de que falantes nativos possuem conhecimento fonético detalhado sobre diferentes níveis (fonológico, morfológico, sintático) de sua língua; ii) O modelo permite a existência de itens prototípicos dentro de uma determinada nuvem de exemplares, no sentido de que um item léxico de uma categoria pode ser um exemplar mais forte ou mais fraco, dependendo, entre outros fatores, de sua freqüência de ocorrência e de seu processamento mais rápido na produção e reconhecimento da fala.

O Modelo de Exemplares fornece explicações plausíveis para a questão da mudança em progresso, sobretudo para o fato de que falantes jovens e idosos costumam ter comportamentos divergentes no que concerne à mudança. Em geral, jovens são mais receptivos a inovações lingüísticas, ao passo que pessoas mais velhas costumam ser mais conservadoras. A justificativa do Modelo de Exemplares para essa divergência de comportamento é a seguinte:

> First, older people may have more exemplars than younger ones for the same pattern, so that the parameter values displayed in older exemplars dominate the production statistics. This line of explanation depends on the assumption that memories decay slowly. A second possibility is that older people are less likely to add new exemplars than young ones; because the formation of new memories becomes less rapid and robust with age, the production statistics are dominated by exemplars stored at a younger age. Differences in attention or in feelings of social affiliation could impact formation of exemplar memories in an analogous way. Both of these lines of explanation predict that the speech patterns of older adults could shift to some extent, just not as rapidly as for younger people. (PIERREHUMBERT, 2001b, p. 149)

Portanto, este modelo assume que a mudança lingüística pode ocorrer porque falantes mais velhos preservam exemplares mais antigos, enquanto falantes mais jovens, na aquisição lingüística, categorizam exemplares mais novos. Ao longo de algumas gerações, esse tipo de comportamento pode causar o enfraquecimento de alguns exemplares e o fortalecimento de outros, gerando, potencialmente, uma mudança lingüística. Com relação a outros correlatos extralingüísticos que também atuam em casos de mudança, Pierrehumbert (2001c) considera que, quando uma regra alofônica entra na língua, existe a probabilidade de esta se aplicar em alguns grupos sociais e não em outros, porque a interação social desempenha papel importante na forma como os seres humanos organizam suas representações mentais de objetos cognoscíveis.

Freqüência tem papel crucial na organização lexical proposta pelo Modelo de Exemplares, já que ela é a responsável pelo fortalecimento ou enfraquecimento dos itens estocados. De forma análoga ao Modelo de Redes, o Modelo de Exemplares pressupõe, também, a existência de duas medidas de freqüência: de tipo ou de ocorrência, basicamente da mesma forma proposta por Bybee (1995, 2001). Com relação ao papel e à importância de cada uma dessas possibilidades de freqüência, Pierrehumbert (2001c) argumenta que as duas categorias estão altamente relacionadas nas línguas do mundo, mas, para a organização dos exemplares no léxico mental, é a freqüência de tipo que desempenha papel mais relevante. Padrões que se repetem em poucas palavras, mesmo que essas tenham alta freqüência de ocorrência, não se espalham para outros itens; em contrapartida, quando um tipo é altamente freqüente na língua (ou seja, é recorrente em muitas palavras), tende a se expandir para outros itens, mesmo que as palavras que compõem essa classe tenham, individualmente, baixa freqüência de ocorrência. Na organização lexical prevista pelo Modelo de Exemplares, a cada vez que um exemplar de uma nuvem é ativado, essa ativação se espalha para os demais itens dessa classe, daí a importância da freqüência de tipo. Por sua vez, a freqüência de ocorrência do item lexical pode concorrer para a

mudança em progresso, da seguinte forma: palavras freqüentes são mais ouvidas do que palavras infreqüentes, então, já que itens freqüentes são mais propensos a mudanças fonéticas redutivas, o falante acaba por ouvir mais ocorrências de palavras que já estão sofrendo um determinado fenômeno fonológico. Dessa forma, "[...] the frequency distribution of outcomes for high frequency words is shifted further in the direction of the historical change than for low frequency words." (PIERREHUMBERT, 2001a, p. 11)

Dentro do Modelo de Exemplares, dois conceitos derivados de freqüência de tipo e de ocorrência são vizinhança lexical ("lexical neighborhood") e fonotática (Cf. PIERREHUMBERT, 2003, p. 189-190). O primeiro conceito diz respeito ao fato de uma palavra se diferenciar de outra em mínimos detalhes, como, por exemplo, a adição, apagamento ou substituição de um único fonema. Itens que diferem entre si apenas por esse tipo de minúcia são alocados em regiões próximas no léxico mental. Por sua vez, a fonotática é a capacidade que o falante tem de julgar se uma seqüência de sons ou até uma palavra inteira é probabilisticamente possível em sua língua nativa. Essa noção de fonotática não é categórica, porque o julgamento sobre uma palavra ou seqüência de sons não diz respeito, em termos absolutos, à existência ou inexistência de tal item ou estrutura na língua. Em vez disso, existe uma gradação entre esses dois julgamentos, havendo palavras e seqüências sonoras mais e menos possíveis dentro de uma determinada língua. Sumarizando, os critérios probabilísticos disponíveis no léxico permitem que o falante julgue certas estruturas como mais ou menos prototípicas na sua língua. Nos extremos desse *continuum* estão a existência e a não-existência de estruturas dentro de uma determinada língua.

Os conceitos de vizinhança lexical e fonotática estão intimamente ligados porque ambos envolvem uma busca no léxico por palavras que apresentam similaridades fonéticas. Com relação aos efeitos de freqüência de que tais conceitos decorrem, Pierrehumbert (2003, p. 188) reporta experimentos demonstrando que o falante leva mais tempo para reconhecer palavras que têm muitos vizinhos lexicais do que para reconhecer palavras que têm poucos vizinhos lexicais. Isso acontece porque, se uma palavra tem muitos vizinhos, seu reconhecimento e recuperação no léxico mental podem ser mais lentos. Além dessa questão de a alta densidade lexical (presença de muitos vizinhos) interferir na velocidade de acesso a uma palavra, "The delay is particularly great if the lexical neighbors have high average frequency relative to the target; the hardest words to recognize are infrequent words with many frequent neighbors, and the easiest are frequent

words with few and infrequent neighbors" (PIERREHUMBERT, 2003, p. 188). Assim, efeitos de freqüência de ocorrência do item lexical atuam no reconhecimento de palavras que são consideradas vizinhas lexicais. Quando a palavra-alvo (aquela que é, realmente, o objeto de produção ou reconhecimento do falante) tem vizinhos de alta freqüência de ocorrência, a concorrência para que essa palavra seja acessada é maior, o que atrasa seu processamento; por outro lado, se a palavra-alvo tiver vizinhos com baixa freqüência de ocorrência, seu reconhecimento é mais rápido porque a concorrência com os vizinhos não é tão grande. Com relação à fonotática, Pierrehumbert (2003, p. 189) também menciona experimentos que mostram que, no julgamento de palavras inventadas, itens similares a outros que já existem são considerados mais possíveis de ocorrerem em uma língua específica. O conceito de fonotática deriva da freqüência de tipo, visto que uma estrutura sonora é considerada mais ou menos possível de acordo com a quantidade de palavras na língua que contém tal seqüência. Assim, quanto maior a freqüência de tipo, maior a probabilidade de determinada estrutura ser possível em uma língua. A interação entre os conceitos de vizinhança lexical e probabilidade fonotática pôde ser observada em experimentos, demonstrando que palavras com fonotática improvável dentro dos padrões de uma língua específica foram consideradas mais aceitas se tivessem mais vizinhos lexicais (PIERREHUMBERT, 2003, p. 190). Tal fato demonstra que o julgamento sobre a existência de uma palavra na língua leva em consideração uma interface entre critérios de vizinhança lexical e fonotática. Convém salientar, no entanto, que, embora correlacionadas, as duas noções têm efeitos contrários na percepção: a boa formação fonotática facilita a percepção, mas a vizinhança lexical muito densa (com muitos exemplares) atrasa a percepção, visto que gera a competição entre diferentes itens.

Conforme afirmado anteriormente, o Modelo de Exemplares difere da teoria gerativa porque postula a gradualidade de categorias lingüísticas e a existência de um léxico organizado em termos de probabilidade, não em termos de possibilidade/impossibilidade de algumas estruturas ocorrerem na língua. Já que o modelo assume que os falantes possuem informações probabilísticas sobre o léxico, cabe a pergunta: especificamente, tais probabilidades são sobre o quê? Pierrehumbert (2000, p. 12) responde a essa questão:

> The ability to form statistical generalizations is recapitulated at higher levels. Phonotactic constraints arise as statistical generalizations over the lexicon, and lexical statistics are gradiently reflected in well-formedness judgments. The knowledge of morphophonological alternations similarly reflects the frequency of word relations in the lexicon.

Percebemos, portanto, que o conhecimento probabilístico perpassa diferentes níveis de conhecimento lingüístico.

Pierrehumbert (2000) discute de forma bastante apropriada a questão dos universais lingüísticos, a partir do que foi proposto pela teoria gerativa:

> [...] languages can differ systematically in arbitrarily fine phonetic detail. This means we do not want to think about universal phonetic categories, but rather about universal phonetic resources which are organized and harnessed by the cognitive system. Water is a physical resource. It is used in different ways in different cultures. [...] Te vowel space – a continuous physical space rendered useful by the connection it establishes between articulation and perception – is also a physical resource. Cultures differ in the way they divide up and use this physical resource. (PIERREHUMBERT, 2000, p. 12)

Percebemos, portanto, que universais não são traços recorrentes em todas as línguas do mundo, mas os recursos físicos disponíveis pela espécie humana para produzir sistemas lingüísticos.

O Modelo de Exemplares apresenta muitos pontos interessantes sobre a categorização de informações lingüísticas através da percepção. No entanto, no que concerne à produção, há algumas lacunas, porque, basicamente, o modelo assume que aquilo que se dá na percepção ocorre ao contrário na produção lingüística. Assim,

> "[...] the decision to produce a given category is realized through activation of that label. The selection of a phonetic target, given the label, may be modeled as a random selection of an exemplar from the cloud of exemplars associated with the label. [...] The likelihood that a particular exemplar will be selected is proportionate to its strength. Production is taken to be sensitive to strength in the same way that perceptual classification is." (PIERREHUMBERT, 2001b, p. 144-145)

A partir dessa citação, percebemos que o modelo carece de algum refinamento teórico que possa analisar melhor as questões pertinentes exclusivamente à produção lingüística, não à percepção da fala. Tal crítica, entretanto, não desqualifica o Modelo de Exemplares, sobretudo por este analisar e acomodar pontos que a teoria gerativa não poderia explicar adequadamente.

Ao longo desta resenha sobre o Modelo de Exemplares, observamos que existem algumas semelhanças importantes entre este e o Modelo de Redes. Em primeiro lugar, ambos admitem a estocagem de informações redundantes (detalhes fonéticos, por exemplo) e de formas já flexionadas (como palavras pluralizadas). No Modelo de Redes, esse tipo de estocagem é importante para a formação de redes a partir de similaridades

fonológicas e semânticas; no Modelo de Exemplares, essa armazenagem garante que membros prototípicos e periféricos sejam determinados. Os dois modelos consideram, também, a atuação das freqüências de tipo e de ocorrência, sendo que ambas as teorias atribuem à freqüência de tipo papel preponderante na formação de classes e generalização de padrões lingüísticos. Finalmente, para os dois modelos, questões relacionadas aos aspectos sociais e pragmáticos das interações comunicativas interferem nas representações mentais de informações lingüísticas. Percebemos, portanto, que muitas características dos dois modelos se sobrepõem, por isso podemos dizer que ambos são complementares, não excludentes.

Conforme mencionamos na primeira parte deste capítulo, o Modelo de Exemplares será, também, adotado como escopo teórico para discutir os casos de plural sob análise nesta tese. A escolha desse modelo se deve ao fato de que ele, assim como o Modelo de Redes, também nega a separação entre léxico e gramática, postulando um léxico plástico e remodelado a partir da experiência do falante com a língua em seu cotidiano.

Com relação às hipóteses do Modelo de Exemplares para os plurais sob discussão nesta tese, podemos dizer que, dadas as similaridades entre este modelo e o de Redes, as hipóteses de ambos seriam análogas. Em linhas gerais, já que a freqüência de tipo é importante para que uma rede de exemplares abarque mais membros, podemos supor que o padrão em *–ões* (mais freqüente para os itens terminados em *–ão* no singular) e o padrão em *–is* (mais freqüente para os itens terminados em *–l* e *–u* no singular) podem se estender a mais exemplares que originariamente pertenciam a outras classes (pluralizados através de *–ãos*, *–ães* ou do acréscimo de *–s*, no caso de itens terminados em ditongo em *–u*). Da mesma forma, se a freqüência de ocorrência faz com que um exemplar se torne mais forte nas representações mentais, pode-se supor que itens altamente freqüentes preservem seus padrões flexionais, ainda que esses sejam irregulares.

Na subseção seguinte, faremos uma breve conclusão sobre os modelos de léxico mental apresentados anteriormente.

4.5 Conclusões

Conforme vimos nas subseções anteriores, existem alguns modelos diferentes de léxico mental, cada um com uma concepção de categorização e estocagem lexical. No entanto, apesar das divergências teóricas, pudemos perceber que tais teorias diferem, principalmente, no papel que cada uma delas confere ao léxico. Nesse sentido, fica claro que a Teoria de Palavras e Regras, por defender a aplicação de regras para gerar formas derivadas, atribui papel reduzido ao léxico, visto que só palavras irregulares estão estocadas. Por outro lado, o Modelo Conexionista, o Modelo de Redes e Modelo de Exemplares conferem importância crucial ao léxico, na medida em que não existem regras preestabelecidas para derivar flexões e as representações mentais são atualizadas a partir da experiência diária dos falantes com a língua. Esses três modelos partilham características análogas sobre o léxico mental. Divergências existentes entre eles já foram abordadas dentro das subseções pertinentes.

Nossa tese tem como objetivo principal analisar os grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u* através das teorias de léxico mental. Conforme já mencionamos e justificamos no Capítulo 1, "Introdução", nossa opção teórica tem como base o Modelo de Redes e o de Exemplares, principalmente por percebermos que a estocagem lexical por eles proposta é consoante com a concepção de plasticidade do léxico mental do falante e com o fato de que o uso interfere diretamente na representação de informações lingüísticas. Apesar da escolha por esses modelos teóricos, neste capítulo intitulado "Quadro Teórico", comentamos, também, o Modelo Conexionista e a Teoria de Palavras e Regras, embora essas não sejam as opções de análise para esta tese. Consideramos que seria importante fazer uma resenha sobre modelos de léxico mental como um todo, a fim de justificar nossa escolha pelo Modelo de Redes e pela Teoria dos Exemplares. No Capítulo 6, "Análise de Dados", procuraremos interpretar os resultados de nossos experimentos à luz dessas duas teorias, discutindo, sobretudo, se um modelo que pressuponha a estocagem por regras seria capaz de explicar a distribuição dos dados coletados.

Com essa decisão em mente, passaremos ao próximo capítulo, que descreve a metodologia elaborada para a coleta dos dados a serem utilizados nesta tese.

CAPÍTULO 5

Metodologia

Iniciamos este capítulo de metodologia retomando alguns pontos já discutidos nos capítulos anteriores, para posteriormente relacionarmos essas informações com a descrição das condutas adotadas em nossa coleta de dados.

Conforme foi afirmado no Capítulo 1, "Introdução", esta tese propõe uma análise dos grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u* do PB. Nesses grupos de palavras, como vimos no Capítulo 2, "Grupos de Plurais", existem regras de pluralização, mas há, também, casos de outras formas de plurais a serem aplicadas. Como exemplos desses itens, podemos citar as seguintes ocorrências, todas registradas em nossa coleta de dados: "órgões" e "escrivões", em vez de "órgãos" e "escrivães"; "difícils" e "sols", no lugar de "difíceis" e "sóis"; "chapéis" e "troféis", em vez de "chapéus" e "troféus".

Alinhando as informações dos Capítulos 2 e 3, "Grupos de Plurais" e "Revisão de Literatura", com as do Capítulo 4, "Quadro Teórico", percebemos que a literatura lingüística do PB não dispõe de uma análise mais precisa sobre essas variações que acontecem nos grupos de plurais terminados em *–ão*, *–l* e ditongo em *–u*. As abordagens mais completas sobre a concordância nominal de número no PB são as de Guy (1981) e Scherre (1988). Tais análises tiveram como escopo a Teoria da Variação e Mudança Lingüística, proposta por Labov (1972) e centraram-se nos processos morfofonológicos de pluralização do PB e nas interações sintáticas entre os itens constituintes de um sintagma nominal. Portanto, essas análises não investigaram fenômenos de pluralização tendo em vista a categorização e o acesso aos itens no léxico mental. Esse não era o objetivo das análises de Guy (1981) e Scherre (1988).

Conforme vimos no Capítulo 4, "Quadro Teórico", as teorias de representação mental analisam correlações entre fenômenos lingüísticos e a forma como o léxico mental está organizado. Na tentativa de entender como as representações lingüísticas estão

estruturadas, questões como freqüência desempenham papel fundamental, pois um dos atributos que a mente humana utiliza para categorizar informações é a freqüência com que certos eventos (em sentido amplo) são vivenciados. Utilizando, basicamente, teorias de categorização, estocagem e acesso ao léxico mental (BYBEE, 1985, 1995, 2001; PIERREHUMBERT, 2000, 2001a, 2001b), esta tese propõe uma nova análise para as classes de plurais terminadas em *–ão*, *–l* e ditongo em *–u* no PB. A inovação que apresentamos é a análise de variações que acontecem nesses grupos (mencionadas no início deste capítulo) a partir do prisma das representações mentais, procurando investigar como o uso diário da língua afeta a forma como o léxico mental categoriza, armazena e acessa informações lingüísticas.

A fim de realizar essa investigação, precisamos de uma metodologia minuciosamente delineada, que nos possibilite captar as peculiaridades envolvidas nesses grupos específicos de plurais do PB (em *–ão*, *–l* e ditongo em *–u*). Por conta disso, os propósitos deste capítulo são: 1) Apresentar a metodologia adotada para conduzir a coleta de dados; 2) Expor dificuldades que os corpora disponíveis para pesquisa lingüística sobre o PB impõem; 3) Enumerar as variáveis lingüísticas e extralingüísticas adotadas em nossa análise, bem como as hipóteses para cada uma delas; 5) Comentar peculiaridades sobre a coleta de dados. Cada uma dessas informações será tratada nas subseções seguintes.

5.1 Metodologia adotada para a coleta de dados

5.1.1 Dados de fala espontânea

A fim de elicitar dados que fossem válidos para a análise dos grupos de plurais em *–ão*, *–l* e ditongo em *–u* do PB e suas possíveis correlações com representações mentais, uma dificuldade inicial se apresentou. Esses grupos de plurais apresentam poucas palavras e, além disso, tais itens têm, em geral, baixa freqüência de ocorrência, sobretudo na língua falada. Como vimos nas Tabelas 5, 8 e 9 do Capítulo 3 ("Revisão de Literatura"), o Corpus ASPA registra 1.389 dados de *–ão*, 679 dados de *–l* e 29 de *–u*, todos nas suas formas

plurais. Essa escassez de itens léxicos para os grupos sob análise foi, portanto, o primeiro problema que enfrentamos na coleta de dados. Nossa intenção inicial era considerar dados de fala espontânea e dados extraídos a partir de experimentos. Tínhamos o intuito de contrapor esses dois resultados, a fim de observar se as generalizações de plural eram mais comuns na fala espontânea ou se também ocorriam em experimentos previamente delineados com o objetivo de fazer o falante pronunciar itens específicos. Para testar a validade dessa metodologia (com entrevistas espontâneas e experimentos psicolingüísticos), realizamos alguns testes-piloto. Esses testes consistiram de entrevistas com falantes e realização de experimentos (conforme descreveremos posteriormente). Com relação às entrevistas, de acordo com Labov (1972), essas foram conduzidas como conversações informais, que focalizaram a história particular de vida do informante, bem como suas opiniões pessoais quanto a questões sociais, históricas e culturais do país e do mundo. Dentro dessas conversas informais, procuramos introduzir perguntas que, de certa forma, induzissem o falante a pronunciar as palavras que eram nosso objeto de estudo. Exemplos dessas perguntas e suas respectivas palavras-alvo (sublinhadas) são: "Você tem <u>irmão</u>?", "Você gosta de ir a <u>museu</u>?" "Você tem o hábito de ler <u>jornal</u>?", dentre outras.

Os experimentos psicolingüísticos foram também utilizados, de modo a verificar se os falantes pronunciavam, a partir desses estímulos, as formas que eram objeto de estudo em nossa tese. Maiores esclarecimentos sobre essa etapa da coleta de dados serão fornecidos posteriormente. Para o presente momento, devemos apenas esclarecer que realizamos alguns testes-piloto com o objetivo de avaliar se poderíamos adotar a entrevista (nos moldes descritos acima) e os experimentos (descritos posteriormente) como fontes para a elicitação dos dados necessários à nossa pesquisa.

Ao todo, nesta etapa de teste-piloto, realizamos três entrevistas com falantes diferentes, selecionados aleatoriamente. Eles não estavam cientes do objetivo da interação, nem de que estávamos fazendo uma pesquisa lingüística. Se os informantes soubessem disso, haveria mais monitoramento e, conseqüentemente, uma fala mais cuidada, o que diminuiria bastante o número de variantes não-padrão, que eram justamente nosso objeto de análise. Sendo assim, dissemos aos informantes que estávamos fazendo uma pesquisa sobre a capacidade de memória das pessoas, procurando identificar se diferentes faixas etárias interfeririam na capacidade de memorização dos indivíduos. Todos os informantes assinaram o "Termo de Consentimento Livre e Esclarecido" (Cf. "Anexo A"), de acordo com a resolução 196/96 do Conselho Nacional de Saúde. As entrevistas duraram em média

30 minutos e foram gravadas utilizando um gravador digital Olympus modelo VN-480PC. Posteriormente, os dados coletados foram digitalizados e codificados de acordo com variáveis lingüísticas e extralingüísticas a serem mencionados nas subseções 5.4 e 5.5 deste capítulo.

Basicamente, nossa avaliação dos testes-piloto realizados concluiu que o número de dados coletados a partir das entrevistas espontâneas era muito pequeno. As dificuldades para a elicitação desses dados eram duas: 1) Em primeiro lugar, os itens terminados em *–ão*, *–l* e ditongo em *–u* são, realmente, pouco freqüentes na língua (conforme estatística apresentada anteriormente), o que, por si só, já dificulta sua ocorrência em conversas informais; 2) Era difícil fazer com que o falante pronunciasse a palavra-alvo no plural. Se perguntávamos, por exemplo: "Você gosta de ler jornal?", com a intenção de que o informante dissesse a palavra no plural, em geral o falante respondia também no singular, como, por exemplo: "Jornal? Ah, não... não leio jornal muito, não". Nos testes-piloto, havia 25 perguntas relacionadas a unidades temáticas em que introduzíamos perguntas com palavras pertencentes aos plurais em *–ão*, *–l* e ditongo em *–u*. Falávamos tais palavras no singular, já que pronunciar a forma no plural induziria a adoção de uma forma específica por parte do falante, mas nosso objetivo era que o informante produzisse, em suas respostas, uma forma espontânea de plural para aqueles itens. Três testes-piloto foram feitos, a fim de verificar se, com essa metodologia, conseguiríamos coletar dados em número suficiente para os plurais em análise nesta tese. Ao fim desses três testes, conseguimos um total de 33 ocorrências de dados válidos, ou seja, itens terminados em *–ão*, *–l* e ditongo em *–u* no plural. Se todos os três falantes tivessem adotado as formas de plural para as 25 palavras que foram apresentadas, haveria 75 dados. As 33 ocorrências dessas palavras correspondem, portanto, a 44% do quantitativo de dados que esperávamos coletar. Com 33 ocorrências extraídas de três informantes, houve uma média de 11 dados por informante, para os três grupos de plurais sob investigação. Consideramos que essa quantidade de ocorrências por falante era insuficiente para que conclusões sobre os plurais analisados pudessem ser esboçadas. Tais problemas dificultaram bastante a coleta de dados válidos para os plurais terminados em *–ão*, *–l* e ditongo em *–u* e fizeram com que buscássemos soluções para que um número maior de dados pudesse ser coletado.

Uma alternativa para elicitar maior quantidade de dados válidos nas entrevistas seria formular as perguntas utilizando as palavras no plural, como, por exemplo: "Você tem irmãos?", "Você gosta de ir a museus?" "Você tem o hábito de ler jornais?". Talvez a

adoção de perguntas com palavras no plural fizesse com que os informantes também utilizassem formas pluralizadas em suas respostas. No entanto, ao adotarmos palavras no plural, de certa forma já estaríamos induzindo o informante a pronunciar as palavras naquela forma de plural que utilizamos. Por exemplo: a palavra "vilão", segundo gramáticas normativas do PB, pode ser pluralizada como "vilões", "vilãos" ou "vilães". Acontece que, se, durante uma entrevista, perguntamos ao informante o que ele acha dos "vilões" das novelas, é pouco provável que esse informante adote, em sua resposta, uma forma de plural diferente daquela que utilizamos em nossa pergunta ("vilões"). Em virtude disso, a opção de formular as perguntas com as palavras no plural não seria viável.

Ainda com relação a essas entrevistas, outra consideração que fizemos foi a seguinte: um dos nossos objetivos era avaliar se a freqüência com que um item léxico é utilizado interfere na sua suscetibilidade a variações de plural. Então, era necessário que o falante pronunciasse palavras de freqüências baixa, média e alta, a fim de que os efeitos de freqüência pudessem ser adequadamente avaliados. No entanto, em fala espontânea, alguns itens infreqüentes são ainda mais difíceis de ocorrer, tais como "mausoléu", "réu", "escrivão", "guardião", "mel" e "sol", para citar dois exemplos de cada um dos grupos analisados. Considerando-se a baixa freqüência de ocorrência desses itens em sua forma plural, como poderíamos extrair esses dados em fala espontânea?

Todos esses motivos nos levaram a considerar que elicitar dados de fala válidos para a análise de freqüência a partir de entrevistas talvez fosse inviável. Apesar disso, não queríamos excluir tais dados (falados), porque sabemos que as variações lingüísticas originam-se na fala para depois alcançarem a modalidade escrita da língua. Assim, começamos a pensar em alternativas possíveis para analisar dados de fala sem realizar entrevistas, já que essas demonstraram não ser eficientes na elicitação dos dados para os plurais em *–ão*, *–l* e ditongo em *–u*.

Uma possibilidade avaliada foi reutilizar os mesmos dados considerados por Scherre (1988) em sua tese de doutorado sobre concordância nominal no PB. Essa amostra é chamada de "Projeto Censo" e é constituída por entrevistas com 64 falantes da cidade do Rio de Janeiro. Tais dados foram coletados no período de 1980 a 1984 com o objetivo de avaliar processos de variação e mudança na fala dos cariocas. O acesso aos dados de tal projeto pode ser feito através da internet, no site <http://www.letras.ufrj.br/~peul/amostras.htm>. Das 64 entrevistas, só conseguimos fazer o *download* de 60, porque quatro delas apresentaram problemas eletrônicos que

impediram o computador de salvar esses arquivos. Nas 60 entrevistas que analisamos, não foi possível contabilizar o número exato de dados, porque as falas do entrevistador e do entrevistando eram transcritas na íntegra, o que inviabilizou a contagem das palavras faladas apenas pelos informantes. O site do "Projeto Censo" não mencionava a quantidade exata de dados desse corpus. Consultamos a tese de Scherre (1988) para verificar se ela mencionava o quantitativo de dados do Projeto Censo, mas ali não havia, também, essa informação.

Apesar de não conhecermos o número exato de dados totais do Projeto Censo, verificamos, em Scherre (1988), a quantidade de ocorrências válidas para os plurais em *–ão* e *–l*, que foram analisados isoladamente pela autora (os itens terminados em ditongo em *–u* foram alocados no grupo de plurais regulares, com acréscimo de *–s* (SCHERRE, 1988)). As Tabelas 6.2.2.22 e 6.2.2.24 daquela tese (Cf. páginas 129 e 132) apresentam, respectivamente, os seguintes quantitativos desses dados no "Projeto Censo": 93 itens terminados em *–l* e 200 pluralizados em *–ães* ou *–ões*[52], codificados em um mesmo grupo. Os itens pluralizados em *–ãos* foram codificados como plural regular e contabilizam 38 dados (Cf. SCHERRE, 1988, p. 125). Apesar de não conhecermos o número total de dados do Projeto Censo, consideramos que a quantidade de dados para os plurais em *–ão* e *–l* que esse corpus apresenta (331, somando-se os três grupos) era muito pequena para que conclusões pudessem ser alcançadas. Por causa disso, descartamos a possibilidade de basear nossa tese nesse corpus.

Considerando-se essas dificuldades para conseguir um número maior de dados de fala para a análise dos grupos de plurais do PB, a decisão que tomamos foi a seguinte: descartamos esses dados (de fala) e elaboramos nossa pesquisa apenas a partir de experimentos lingüísticos delineados com o objetivo de fazer o informante pronunciar as palavras que eram objeto de estudo em nossa tese. É claro que, idealmente, seria interessante realizar comparações entre os dados extraídos desses experimentos e dados de fala; no entanto, por considerarmos, principalmente, a baixa ocorrência de itens terminados em *–ão*, *–l* e ditongo em *–u*, resolvemos analisar dados oriundos apenas de experimentos.

---

[52] Os itens em *–ão* foram codificados por Scherre (1988) da seguinte maneira: palavras pluralizadas em *–ões* e *–ães* ("leões", "pães", por exemplo) foram consideradas como pertencentes à formação de plural em *–ão*; palavras com plural em *–ãos* foram codificadas como regulares, pluralizadas a partir da adição do morfema *–s* (Cf. Capítulo 3, "Revisão de Literatura"). Os 200 dados a que nos referimos aqui, são, portanto, relativos ao grupo que se pluraliza com os morfemas *–ões* e *–ães*. Os demais itens são contabilizados por Scherre (1988) juntamente com as palavras de pluralização regular.

Devemos salientar que estamos cientes de que a opção de trabalhar somente com experimentos lingüísticos apresenta problemas. Em primeiro lugar, na realização de testes dessa natureza, perdemos a extração do vernáculo, modalidade de fala em que formas não-padrão ocorrem de forma espontânea. Além disso, os falantes costumam apresentar algum nervosismo quando sabem que estão sendo submetidos a uma situação de teste, o que, em teoria, poderia afetar o desempenho de nossos informantes e fazer com que, nos experimentos, houvesse respostas que não ocorrem na fala espontânea. No entanto, apesar dessas limitações dos experimentos psicolingüísticos, decidimos optar por tal modalidade de elicitação de dados, por dois motivos principais, a saber: 1) A ocorrência de itens terminados em *–ão*, *–l* e ditongo em *–u* em situações de fala espontânea é muito baixa (Cf. freqüência no ASPA e Projeto Censo); além disso, um dos nossos objetivos era avaliar se a freqüência de ocorrência do item léxico interferiria na possível generalização de plurais. Dessa forma, precisávamos coletar dados de faixas diferentes de plural (baixa, média e alta). Alguns itens de alta freqüência ocorrem com mais facilidade em entrevistas espontâneas, como "jornal", "museu" e "irmão". No entanto, itens de baixa freqüência, como "escrivão", "jirau", "sol", dificilmente seriam elicitados em entrevistas, ainda mais em suas formas plurais, por isso seria inviável analisar efeitos de freqüência a partir de entrevistas espontâneas; 2) Em segundo lugar, o Modelo de Redes e a Teoria dos Exemplares (adotados como escopo teórico para nossa tese) utilizam, em algumas de suas análises, dados extraídos apenas a partir de experimentos. Como exemplos disso, podemos citar: i) Bybee e Moder (1983), que elaboraram um experimento com palavras hipotéticas com o intuito de analisar como os falantes flexionariam verbos inventados que partilhassem semelhanças com os verbos irregulares do paradigma "string"/"strung" no inglês; ii) Bybee e Newman (1995), que criaram uma língua hipotética a fim de observar se os falantes, flexionando as palavras dessa língua, adotariam, em maior proporção, sufixação ou mudança de radical; iii) Treiman *et al.* (2000) *apud* Pierrehumbert (2001c), que organizaram experimentos para observar a sensibilidade dos falantes a restrições fonotáticas da língua. Observamos, portanto, que a elaboração de experimentos é uma prática usualmente adotada pelo Modelo de Redes e pela Teoria dos Exemplares. Por causa desses motivos, consideramos que utilizar somente testes psicolingüísticos não invalidaria as conclusões alcançadas por nossa tese.

A subseção seguinte descreve como foram elaborados e aplicados os experimentos utilizados para a elicitação de dados em nossa pesquisa.

5.1.2 Dados de fala não-espontânea

A partir do que descrevemos na subseção anterior sobre a dificuldade de ocorrência de palavras terminadas em *–ão*, *–l* e ditongo em *–u* em fala espontânea, elaboramos três modalidades diferentes de experimentos para extrair dados válidos para esses grupos de plurais. Explicamos, abaixo, cada uma dessas modalidades, exatamente na ordem em que foram apresentadas aos falantes.

1. Conjunto de Figuras

A primeira modalidade de coleta de dados foi feita através de gravuras. Escolhemos, na internet ou em revistas, algumas figuras de palavras pré-selecionadas segundo critérios específicos (mencionados a seguir) e fizemos um conjunto de figuras com elas. Como estávamos analisando a pluralização das palavras, selecionamos gravuras em que os elementos representados aparecessem mais de uma vez, como "vulcões" em erupção, uma propaganda com vários "anéis" ou uma fotografia com um móvel que sustentava vários "chapéus". Quanto à escolha das palavras utilizadas nessa etapa da coleta de dados, só pudemos utilizar substantivos concretos, visto que adjetivos e substantivos abstratos, como "escrivão", "europeu", "responsável", não apresentam representação gráfica facilmente dedutível. A elicitação desse tipo de palavra ficou restrita à última etapa da coleta dos dados, a ser descrita posteriormente. A existência de palavras terminadas em *–ão*, *–l* e ditongo em *–u* é tão pequena (Cf. Tabelas 5, 8 e 9 do Capítulo 3, "Revisão de Literatura"), que tivemos de nos limitar a utilizar, nesta fase da coleta de dados, os itens para os quais conseguimos encontrar figuras. Por causa da baixa freqüência desses itens, não houve critérios fonológicos, morfológicos ou semânticos que pudessem ser utilizados na seleção das gravuras. O único parâmetro com o qual pudemos contar foi a freqüência de ocorrência da palavra. Como salientamos no Capítulo 4, "Quadro Teórico", as teorias que adotamos nesta tese (Modelo de Redes e Teoria dos Exemplares) conferem à freqüência (de ocorrência ou de tipo) papel crucial na organização das representações lingüísticas no léxico mental dos falantes. Além disso, a interação entre tais freqüências é apontada como um dos fatores responsáveis pelo remodelamento da gramática enquanto a língua é usada no cotidiano (Cf. HOPPER, 1998). Selecionamos, então, palavras utilizando o critério

freqüência de ocorrência, a fim de verificar se o fato de a palavra ser mais ou menos recorrente interferiria na sua suscetibilidade a variações de plural.

Ainda quanto à seleção de palavras, outro detalhe precisa ser esclarecido. A fim de evitar que os falantes percebessem exatamente quais eram as palavras sob investigação, escolhemos, também, figuras para "distrair" a atenção dos informantes e evitar que eles descobrissem os grupos de plurais que estavam sendo analisados. Selecionamos, então, seis palavras "distratoras" (Cf. Tabela 12 abaixo) e misturamos essas figuras àquelas que representavam as palavras consideradas em nossa análise. A escolha dessas palavras também teve como critério principal o fator freqüência de ocorrência, pelos mesmos motivos citados anteriormente para os demais grupos de palavras. Sendo assim, utilizamos 12 palavras terminadas em *–ão*[53], nove terminadas em *–l*, seis terminadas em ditongo em *–u* e seis distratoras. A diferença no quantitativo de palavras de cada um dos grupos ocorreu, sobretudo, devido à dificuldade de encontrar itens léxicos dessas classes de plural que pudessem ser graficamente representados. Especialmente no grupo de palavras terminadas em ditongo em *–u*, a pequena quantidade de itens (Cf. freqüência ASPA e Novo Dicionário Aurélio Eletrônico, nas Tabelas 9 e 11) impediu a utilização de mais que seis palavras. Esse experimento, tal como foi apresentado aos informantes, é reproduzido no "Anexo A".

As gravuras selecionadas foram impressas, recortadas e reunidas aleatoriamente em um conjunto de figuras. Apresentamos as figuras ao falante, uma a uma, na mesma ordem em todas as entrevistas, e pedimos que o informante fizesse um comentário sobre o que via. Organizamos, abaixo, uma tabela com as 33 palavras utilizadas no "Conjunto de Figuras" e suas freqüências de ocorrência no plural. Os itens são apresentados por grupo de plural analisado, de acordo com a freqüência no plural (em ordem crescente). As freqüências foram consultadas no Corpus NILC/São Carlos[54].

---

[53] Os itens terminados em *–ão* têm três possibilidades de plural, por isso incluímos um número maior de itens dessa classe no experimento (12 palavras).

[54] Mais detalhes sobre o Corpus NILC/São Carlos serão fornecidos na subseção 5.2.1 deste capítulo.

Tabela 12: Palavras selecionadas para o "Conjunto de Figuras" (freqüência plural)

Palavras terminadas em *–ão*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
pião[55]	1	balão	110	caminhão	742
limão	13	pão	118	avião	874
escorpião	23	pulmão	164	cartão	941
vulcão	62	grão	471	mão	2.949
Palavras terminadas em *–l*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
cachecol	2	farol	100	móvel	728
avental	17	anel	110	sinal	1.422
pincel	37	míssil	237	jornal	2.490
Palavras terminadas em ditongo em *–u*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
berimbau	0	degrau	109	pneu	554
troféu	72	chapéu	125	grau	916
Palavras distratoras					
Freqüência Baixa (0-500)		**Freqüência Média (501-1.000)**		**Freqüência Alta (+ de 1.000)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
gato	270	flor	840	quadro	1.069
relógio	286	árvore	845	estrela	1.555

Com relação às palavras listadas na tabela acima, algumas explicações são necessárias:

1. Conforme Bybee (2006) afirma, não existem, ainda, medidas precisas para definir exatamente o que são, em termos absolutos, faixas de freqüência alta, média e baixa. Segundo a autora, essa é uma questão empírica e valores categóricos de freqüência para cada fenômeno lingüístico serão definidos conforme as análises lingüísticas evoluírem. Em função disso, não podemos contar, até o presente momento, com limiares precisos para definir as faixas de freqüência para os itens

[55] Para as palavras que, segundo gramáticas do PB (Cf. subseção 2.2.2), apresentavam plural abundante, consultamos as duas ou três possibilidades de plural existentes. Os números da tabela acima representam, portanto, o total da freqüência das duas ou três formas para os casos de plurais abundantes.

sob análise em nossa tese. Como a investigação de efeitos de freqüência era um dos principais focos de nosso estudo, tivemos de definir três faixas de freqüência para os itens terminados em *–ão*, *–l* e ditongo em *–u*, ainda que essa fosse uma dificuldade apontada pelas teorias adotadas em nossa pesquisa[56];

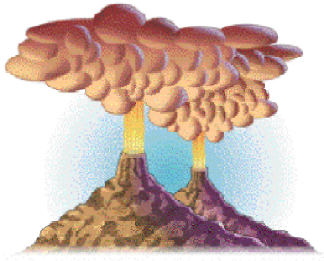
2. No caso das palavras "distratoras", as faixas de freqüência são mais amplas porque todos os itens (com exceção de "flor") são pluralizados a partir da adição de morfema *–s*. Essa é a regra geral de formação de plural no PB e abarca, portanto, a maior quantidade de palavras da nossa língua, por isso optamos por faixas de freqüência mais abrangentes (freqüência baixa entre zero e 500 ocorrências, média entre 501 e 1.000, e alta com mais de 1.000 ocorrências);
3. Com relação aos itens terminados em *–ão*, nossa intenção inicial era utilizar, nas diferentes faixas de freqüência, palavras pluralizadas etimologicamente com os três morfemas (*–ões*, *–ãos* e *–ães*). No entanto, a única palavra do PB pluralizada em *–ães* e que apresenta mais de 500 ocorrências (freqüência alta), segundo o Corpus NILC/São Carlos, é "alemães", cuja representação gráfica é difícil. Por conta desse tipo de lacuna no léxico, não foi possível alocar palavras com diferentes pluralizações em cada uma das faixas de freqüência. Talvez justamente por causa da existência dessas lacunas no léxico generalizações de plural estejam ocorrendo. Se itens pluralizados em *–ãos* e *–ães* fossem muito freqüentes em suas formas flexionadas, as variações de plural nessas classes seriam casos esporádicos. Conforme veremos em nosso Capítulo 6, "Análise de Dados", a adoção de plurais analógicos está bastante avançada no grupo de palavras com plurais etimológicos em *–ãos* e *–ães*, possivelmente por causa dessas lacunas na quantidade de itens flexionados com esses morfemas.

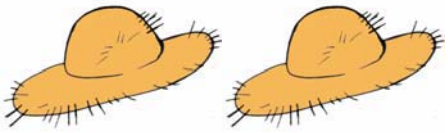
Feitos os esclarecimentos necessários ao entendimento desta primeira etapa da coleta de dados, explicaremos o segundo experimento utilizado em nossa tese.

[56] Na subseção 5.3 deste capítulo, retomaremos essa discussão sobre freqüência, justificando como fizemos a divisão das faixas diferentes freqüência para os grupos de plurais analisados nesta tese.

2. Leitura de Frases

Esta segunda modalidade da coleta de dados consistiu na leitura de algumas frases. No meio dessas sentenças, colocamos figuras de palavras terminadas em *–ão*, *–l* ou ditongo em *–u*. Essas gravuras representavam objetos sempre em número maior que um, de modo a induzir o informante a adotar uma forma de plural para a figura que via. Seguem abaixo dois exemplos dessa leitura:

Os  da Califórnia sempre entram em erupção.

Ainda estamos precisando de dois  de palha emprestados para a quadrilha.

Fazendo a leitura das frases acima, o falante, obrigatoriamente, teve de optar por uma das formas de pluralização das palavras "vulcão" e "chapéu": "vulc<u>ões</u>", "vulc<u>ãos</u>" ou "vulc<u>ães</u>" e "chapé<u>us</u>" ou "chapé<u>is</u>", respectivamente. É claro que ele poderia, também, escolher a opção singular ("Os vulcãoØ da Califórnia" ou "dois chapéuØ de palha") ou adotar uma marca diferente dos morfemas de plural ("vulcõe" ou "chapéi", por exemplo). No entanto, a ocorrência de tais dados (no singular ou com outro plural) foi pequena em nosso corpus (Cf. Tabelas 20, 34 e 45, no Capítulo 6, "Análise de Dados"). Na maioria dos casos, os falantes optaram, realmente, por uma forma de plural para os itens representados através das figuras.

Para a seleção das figuras utilizadas na "Leitura de Frases", deparamo-nos com os mesmos problemas encontrados na escolha das gravuras para o experimento anterior ("Conjunto de Figuras"). Como o número de palavras pertencentes aos grupos terminados em *–ão*, *–l* e ditongo em *–u* é pequeno e precisávamos utilizar apenas palavras que fossem graficamente representáveis, ficamos com uma quantidade reduzida de possibilidades de itens para inserir nesta etapa dos experimentos. Por causa disso, algumas palavras utilizadas no primeiro experimento foram repetidas neste segundo. A elicitação de palavras abstratas, como "europeu", "guardião", "infantil", ficou restrita apenas à última etapa dos experimentos, conforme veremos a seguir. Quanto aos critérios para selecionar as palavras utilizadas na "Leitura de Frases", contamos, também, com a freqüência de ocorrência, nos mesmos moldes e faixas considerados para o "Conjunto de Figuras". Palavras distratoras foram, também, utilizadas, de modo a fazer com que os itens que eram objeto de investigação na tese não fossem facilmente percebidos pelos falantes[57].

Segue abaixo a lista de palavras utilizadas na "Leitura de Frases". Os itens estão organizados por grupo de pluralização e freqüência de ocorrência no plural, em ordem crescente. As medidas de freqüência também foram feitas no Corpus NILC/São Carlos, para o qual mencionaremos mais detalhes posteriormente, ainda neste capítulo. Uma lista completa com todas as frases apresentadas aos informantes encontra-se no "Anexo A". Finalmente, apresentamos, abaixo, a tabela com as 30 palavras utilizadas neste experimento:

[57] As palavras distratoras adotadas nesta modalidade foram diferentes daquelas utilizadas no experimento anterior.

Tabela 13: Palavras selecionadas para a "Leitura de Frases" (freqüência plural)

Palavras terminadas em *–ão*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
pião	1	pão	118	caminhão	792
pavão	5	televisão	122	avião	874
vulcão	62	leão	125	mão	2.949
Palavras terminadas em *–l*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
funil	2	farol	100	móvel	728
sol	27	anel	110	sinal	1.422
pastel	63	míssil	237	jornal	2.490
Palavras terminadas em ditongo em *–u*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
berimbau	0	degrau	109	pneu	554
troféu	72	chapéu	125	grau	916
Palavras distratoras					
Freqüência Baixa (0-500)		**Freqüência Média (501-1.000)**		**Freqüência Alta (+ de 1.000)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
borboleta	65	ovo	531	pé	2.970
cachorro	182	peixe	583	livro	3.230

Conforme a tabela acima demonstra, as palavras selecionadas para a "Leitura de Frases" compõem uma amostra mais homogênea, de modo que os itens terminados em *–ão* e *–l* apresentam exatamente a mesma quantidade de palavras (nove itens para cada), divididas nas faixas de freqüência. Nos itens terminados em *–u*, mais uma vez as limitações do léxico e a dificuldade de representação gráfica não nos permitiram alocar três itens em cada uma das faixas de freqüência. Itens como "europeu" e "judeu" apresentam, no Corpus NILC/São Carlos, freqüência acima de 500 ocorrências, o que as classificaria nas células de freqüência alta em nossa tabela. No entanto, não seria possível encontrar figuras que pudessem representar essas palavras. Por causa desse problema, tivemos de limitar o número de dados para seis nesse grupo de palavras.

Descrevemos, a seguir, a última etapa da coleta dos dados.

3. Teste de Reação

Basicamente, este teste foi organizado da seguinte maneira: elaboramos uma lista de palavras pertencentes aos grupos de plurais analisados em nossa tese, acrescentamos alguns itens distratores (12, no total, conforme mostra a Tabela 14 abaixo), e perguntamos aos informantes o plural de cada um dos itens escolhidos. Sendo assim, a pesquisadora falava a palavra no singular e a tarefa do informante era prover o plural desse item.

Inicialmente, nossa idéia para este teste era medir o tempo de reação do falante para falar a forma plural de um item singular apresentado pela pesquisadora. Os dados seriam digitalizados e mediríamos, em programas de análise acústica, o tempo de reação dos falantes para cada uma das palavras, a fim de verificar se haveria respostas mais rápidas ou mais lentas e que fatores estariam interferindo nessa velocidade. No entanto, para que tal análise fosse feita, os dados deveriam ser de uma qualidade acústica muito boa, o que só poderia ser alcançado se a gravação tivesse sido feita em uma cabine acústica, com equipamento de alta qualidade. Como fizemos as entrevistas no local de trabalho ou na casa dos informantes, a existência de ruídos externos inerentes ao meio ambiente impossibilitou a medida do tempo de reação dos falantes. Por causa desses ruídos, era impossível precisar onde terminava a palavra pronunciada pela pesquisadora e onde começava a resposta dita pelo informante. Sendo assim, tivemos de buscar outras possibilidades de utilização de tais dados, já que a análise acústica era inviável[58].

Coletados os dados, observamos que, em alguns casos, os falantes apresentavam hesitações sobre os plurais de certas palavras. Por isso, decidimos avaliar, neste experimento, não apenas as formas que os falantes adotaram como plurais às palavras apresentadas, mas também a maneira como eles reagiram quando proveram as respostas. Nossa hipótese para este experimento era de que as palavras menos freqüentes teriam acesso mais lento no léxico mental e, conseqüentemente, o falante hesitaria mais para pronunciar o plural de uma palavra infreqüente do que o de uma palavra freqüente. Na realização deste experimento, observamos que algumas respostas eram prontamente dadas, sem que fosse necessário o falante parar para pensar na forma de plural adequada; no entanto, em outras, os falantes demonstravam dúvidas, o que gerou hesitações de dois tipos: i) **Silenciosa**, em que o falante parava para pensar sobre o plural da palavra e ficava

---

[58] Agradecemos muito à Professora Eleonora Albano por suas valiosas contribuições na elaboração da metodologia e das possibilidades de análise deste experimento.

em silêncio no intervalo entre o item pronunciado pela pesquisadora e a forma de plural adotada pelo informante; e ii) **Preenchida**, em que o falante também demorava para prover a resposta, mas, em vez de ficar em silêncio no intervalo entre as duas palavras (da pesquisadora e resposta), fazia algum tipo de som. Nesse segundo caso, encontramos os seguintes tipos de hesitações:

- Murmúrio antes da resposta ("Uuuuhhhhhh...", "Eeeeehhhhh...");
- Questionamento sobre a forma de plural daquela palavra ("Plural de mel?");
- Gaguejo no início da palavra ("CCCCéus");
- Alongamento de vogal antes do morfema de plural ("Degraaaaaaau");
- Mudança de resposta ("Degrais. Não, degrais, não, degraus!");
- Utilização de outra palavra antes da resposta ("São paus");
- Repetição da palavra no singular dita pela pesquisadora ("Sol?");
- Pronúncia pausada da palavra ("Mau-so-léus");
- Interrupção da palavra na metade ("Crist – cristões").

Como nesta etapa da coleta de dados não dependíamos de figuras que representassem graficamente as palavras, pudemos selecionar itens abstratos, como "cristão", "judeu" e "responsável". Tais dados haviam sido excluídos das etapas anteriores dos experimentos porque havia, naqueles testes, necessidade de que a palavra fosse graficamente representável.

A freqüência de ocorrência do item lexical foi, também, um dos fatores que determinou a escolha das palavras para o "Teste de Reação". A fonte dos dados sobre freqüência também foi o Corpus NILC/São Carlos, da mesma maneira que para as demais etapas dos experimentos. A tabela abaixo apresenta, em ordem crescente de freqüência, as 65 palavras selecionadas para esta etapa da coleta de dados:

Tabela 14: Palavras selecionadas para o "Teste de Reação" (freqüência plural)

Palavras terminadas em *–ão*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
limão	13	pão	118	opinião	571
escrivão	14	anão	142	avião	874
bênção	26	religião	191	alemão	1.025
guardião	26	cristão	235	irmão	1.125
união	34	cão	350	órgão	1.578
vulcão	62	exposição	494	mão	2.949
Palavras terminadas em *–l*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
mel	0	anel	110	espanhol	501
anzol	8	lençol	118	difícil	654
avental	17	sal	129	útil	783
sol	27	acessível	165	hospital	1.513
gentil	31	azul	280	responsável	1.894
agradável	74	infantil	328	gol	4.565
Palavras terminadas em ditongo em *–u*					
Freqüência Baixa (0-99)		**Freqüência Média (100-500)**		**Freqüência Alta (+ de 500)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
jirau	1	degrau	109	judeu	808
mausoléu	4	chapéu	125	grau	916
ateu	24	céu	163	europeu	1.119
véu	33	réu	309	meu	2.139
pau	46	mau	346	seu	29.682
troféu	72	museu	424		
Palavras distratoras					
Freqüência Baixa (0-500)		**Freqüência Média (501-1.000)**		**Freqüência Alta (+ de 1.000)**	
Palavras	**Ocorrências**	**Palavras**	**Ocorrências**	**Palavras**	**Ocorrências**
girafa	11	vizinho	738	mercado	1.195
batata	157	árvore	845	antigo	1.255
riqueza	171	menino	938	humano	2.112
barato	398	domingo	980	alimento	2.595

Nos grupos terminados em *–ão* e *–l*, conseguimos alocar, de forma homogênea, seis palavras em cada faixa de freqüência. No entanto, no grupo terminado em

ditongo em *–u*, não havia seis palavras que apresentassem freqüência de ocorrência maior que 500, por isso nossa freqüência alta para essa classe conta com apenas cinco itens. Essa lacuna é uma característica do léxico que se reflete na seleção dos dados para nossa pesquisa. Finalmente, quanto às palavras distratoras, foram escolhidos quatro itens para cada faixa de freqüência. Como o número de palavras para o "Teste de Reação" era maior que o das demais etapas de coleta, resolvemos utilizar, também, um número maior de palavras distratoras. Sobre essas palavras, tomamos, ainda, outra decisão. Em experimentos-piloto que realizamos, observamos que, no "Teste de Reação", os falantes demonstravam certo nervosismo quando eram informados de que deveriam prover o plural de uma lista de palavras. Devido a essa apreensão, alguns falantes tenderam a começar o teste pronunciando respostas muito rápidas, diminuindo a velocidade à medida que a lista de palavras prosseguia e os informantes se sentiam mais calmos. A conseqüência disso foi que as primeiras palavras do teste tinham respostas mais rápidas do que as últimas, e isso não era em decorrência de fatores lingüísticos, mas da insegurança inicial dos falantes. A solução que encontramos para minimizar esse problema foi a seguinte: começamos o "Teste de Reação" utilizando apenas palavras distratoras, já que essas não eram o foco específico de nossa investigação e sua velocidade de acesso não estava sendo analisada. Assim, o teste continha, inicialmente, oito palavras distratoras, e, após esses itens, as palavras terminadas em *–ão*, *–l* e ditongo em *–u* eram intercaladas com as distratoras restantes, que ainda não haviam sido utilizadas no início do teste. Com essa estratégia, observamos que, ao iniciarmos a extração dos dados realmente válidos (a partir da nona palavra), os falantes já estavam mais calmos e a velocidade na resposta à palavra poderia ser considerada independentemente da interferência de fatores psicológicos. O teste foi realizado individualmente e de uma só vez, não havendo intervalo para descanso dos informantes. O "Anexo A" mostra o "Teste de Reação" com as palavras na ordem exata em que foram apresentadas aos falantes.

Basicamente, essas foram as três formas de elicitação de dados utilizadas na metodologia desta tese. Maiores detalhes sobre a realização do trabalho de campo serão mencionados na subseção 5.6 deste capítulo.

Retomando os critérios para a seleção das palavras utilizadas nos experimentos, o principal parâmetro considerado foi freqüência de ocorrência do item léxico. Procuramos diversificar a estrutura dos itens selecionados, levando em conta fatores como estrutura morfológica, número de sílabas e tonicidade; no entanto, como havia a dificuldade de

representação gráfica, tínhamos de optar por um critério básico para a seleção de palavras, mesmo sem perder de vista outros parâmetros subsidiários que deveriam ser considerados também, sobretudo na codificação dos dados (Cf. subseção 5.4 neste capítulo).

A escolha do fator freqüência de ocorrência do item lexical como critério básico ocorreu porque, conforme vimos no Capítulo 4, "Quadro Teórico", esse ponto é, para o Modelo de Redes e para a Teoria dos Exemplares, um dos mais importantes na categorização e acesso às palavras no léxico mental. A literatura sobre representações mentais indica a freqüência como um dos pontos determinantes para que variações lingüísticas sejam desencadeadas. Por causa disso, definimos a freqüência como parâmetro básico de escolha das palavras a serem utilizadas nos experimentos descritos, mas essa opção nos trouxe algumas dificuldades que acarretaram decisões metodológicas necessárias. O primeiro problema relacionado à contagem de freqüência de ocorrência diz respeito à adoção de um corpus representativo do PB no qual essa busca por itens léxicos pudesse ser feita. O PB dispõe de alguns corpora para análise lingüística, no entanto cada um deles tem propriedades específicas e foi necessário considerar com cautela qual deles seria adequado para a investigação proposta nesta tese. Por isso, na subseção seguinte, descrevemos os corpora do PB disponíveis para pesquisas lingüísticas e justificamos nossa adoção do Corpus NILC/São Carlos para a consulta das freqüências de ocorrência das palavras utilizadas em nossos experimentos.

5.2 Corpora do PB consultados

Segundo o Modelo de Redes e a Teoria dos Exemplares, qualquer pesquisa que tenha como objetivo avaliar como acontecem a categorização e a estocagem das informações lingüísticas na mente do falante precisa considerar medidas de freqüência de tipo e de ocorrência (Cf. Capítulo 4, "Quadro Teórico"). Naturalmente, para proceder a análises refinadas dessa natureza, é necessário contar com corpora lingüísticos que propiciem opções de busca detalhadas.

Basicamente, os corpora consultados em nossa tese foram três: o Corpus NILC/Universidade de São Carlos, o Corpus LAEL (Lingüística Aplicada e Estudos da

Linguagem), e o Corpus ASPA (Avaliação Sonora do Português Atual). Nas subseções seguintes, vamos comentar esses corpora individualmente.

5.2.1 O Corpus NILC/São Carlos

O Corpus NILC/São Carlos (disponível em <www.linguateca.pt/ACDC/>) é parte do Corpus CETEM/Público (Corpus de Extratos de Textos Eletrônicos MCT/ Público). Na tabela abaixo, organizamos os dados quantitativos desse corpus e suas fontes.

Tabela 15: Dados sobre o Corpus NILC/São Carlos

Corpus SÃO CARLOS	Número de formas	Número de tipos	Tipo	Descrição	Tamanho
Unidades	41.372.943	457.556	DI	texto didáctico	423.893
Total de Palavras	32.091.996	433.030	ENC	enciclopédia	283.838
Palavras em minúscula	23.217.976	158.261	ENS	ensaio	2.177.193
Palavras com inicial maiúscula	4.595.495	127.721	EP	texto epistolar	3.338
Palavras todas em maiúsculas	464.974	22.996	JO	jornalístico	29.462.874
Números	427.186	2.978	JOCF	jornalístico só CETENFolha	29.462.874
Palavras com números	38.568	4.379	JO	jornalístico sem CETENFolha	29.462.874
Palavras mistas	98.344	9.267	LE	texto legal	1.083.200
Pontuação	2.415.505	24.522	LI	literário	919.628
			RE	revista	153.454

Fonte: <http://www.linguateca.pt/CETEMPublico/>

Observamos, portanto, que o Corpus NILC/São Carlos apresenta uma grande quantidade de ocorrências (41.372.943 dados), mas tais itens são extraídos somente de textos escritos. Como as variações lingüísticas originam-se na fala, o ideal seria que pudéssemos, em nossa pesquisa, consultar um corpus de língua falada. Essa opção não está disponível no Corpus NILC/São Carlos, já que ele é formado apenas por itens de língua escrita. Por outro lado, a tabela acima demonstra que esse corpus apresenta uma grande

diversidade de tipos de texto, o que faz com que ele seja bastante representativo da língua escrita.

Basicamente, o Corpus NILC/São Carlos provê uma busca por ocorrências. Então, por exemplo, se estamos interessados em saber a freqüência de ocorrência da palavra "milhão", é só digitar a palavra e, em alguns segundos, o site apresentará os resultados. Esse corpus propicia a busca por palavras isoladas, como "céu", "lençol", "milhão", ou por itens em contexto, como "céu azul", "maus lençóis" ou "um milhão". No entanto, uma característica que tal corpus apresenta é o fato de fazer distinções entre letras maiúsculas, minúsculas e marcas de acentuação, e não existir uma opção que nos permita ignorar tais detalhes. Sendo assim, se queremos saber a real ocorrência da palavra "milhões", é preciso procurá-la de, no mínimo, seis maneiras diferentes:

milhões – 20.110 ocorrências
MILHÕES – 15 ocorrências
Milhões– 48 ocorrências
milhoes – 5 ocorrências
MILHOES – 0 ocorrências
Milhoes – 0 ocorrências

No caso das generalizações de plural em questão, em que estamos lidando com palavras terminadas em *–ão*, muitos desses itens apresentam o sufixo *–ação*, que contém, além do til, "c" com cedilha. Isso multiplica o trabalho de busca no Corpus NILC/São Carlos porque, se quisermos uma busca realmente minuciosa, precisamos procurar a palavra com todas as possibilidades de grafia, uma vez que a pessoa que digitou a palavra que está no corpus pode ter cometido algum erro, deixando de acrescentar a cedilha ou o til ou até mesmo trocando tais marcas. Como nossa tese tem por objetivo analisar variações nos grupos de plural, e não variações na forma ortográfica das palavras, adotamos a busca da palavra em sua forma minúscula, com os acentos necessários segundo a ortografia vigente, quando fosse o caso. Após compararmos diferentes palavras dos grupos de plural sob análise e as ocorrências para cada uma das grafias possíveis, concluímos que uma busca mais simplificada não prejudicaria os objetivos primários de nossa pesquisa.

5.2.2 O Corpus LAEL

O Corpus LAEL (Lingüística Aplicada e Estudos da Linguagem, disponível em <http://lael.pucsp.br/corpora/index.htm>) é dividido em dados de fala e de escrita. Sua composição é a seguinte:

Tabela 16: Totais de palavras dos registros e tipos de textos do Corpus LAEL

Registros de texto	Total de palavras de cada registro	Tipos de texto	Total de palavras de cada tipo de texto
1. Registro acadêmico	198.652	1) Artigos e teses acadêmicas	198.652
		1) Cartas comerciais	19.736
		2) Cartas de pedido de emprego	14.306
2. Registros de negócios, comerciais e técnicos	386.138	3) Editais	30.829
		4) Fax Comerciais	16.131
		5) Relatórios anuais de negócio	104.255
		6) Manuais de informática	200.881
3. Registros falados	197.901	1) Aulas	84.910
		2) Conversação	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) *apud* FONTES MARTINS, 2007, a ser publicado.

Uma suposta vantagem do Corpus LAEL[59] sobre o NILC/São Carlos é o fato de o primeiro apresentar dados de fala, enquanto o segundo disponibiliza apenas dados de língua escrita. No entanto, a partir dos números da tabela acima, observamos que os dados de fala (aulas e conversações) do LAEL representam 197.901 dados, ou seja, 16,7% do total do corpus. Sendo assim, o número de itens de fala considerados nesse corpus não é muito expressivo, visto que a maioria dos dados (83,3%) é oriunda da escrita.

Uma desvantagem desse corpus é que a transcrição dos dados, mesmo os relativos à fala, é feita ortograficamente, o que impede, por exemplo, a busca por sutilezas

[59] Nesta tese, o Corpus LAEL a que nos referimos é aquele disponibilizado on-line, no endereço supracitado. Não fizemos consulta diretamente ao Corpus LAEL na PUC de São Paulo.

fonéticas como plurais terminados em [ɛws] ou [ews]. Sendo assim, a suposta vantagem do Corpus LAEL, por utilizar dados de fala, não é tão prática, visto que a modalidade falada compõe uma parte pequena do corpus (16,7%) e peculiaridades inerentes à pronúncia das palavras não podem ser consultadas.

Uma característica importante do Corpus LAEL (tanto de fala quanto de escrita) é que ele provê uma lista de freqüência de ocorrência dos itens léxicos, que elenca todas as palavras do corpus, da mais freqüente à menos freqüente. Essa organização dos dados viabiliza pesquisas como a listagem de palavras mais freqüentes em cada um dos grupos de plurais sob análise nesta tese. O NILC/São Carlos oferece apenas a busca de freqüência de ocorrência da palavra, não disponibilizando, portanto, tal lista dos itens mais freqüentes aos menos freqüentes.

Da mesma forma que o Corpus NILC/São Carlos, o LAEL também faz distinção de acentuação. Assim, a mesma opção feita para a coleta de dados no NILC/São Carlos foi adotada para o LAEL, ou seja, procuramos a palavra com sua ortografia oficial e em caracteres minúsculos. Já comprovamos anteriormente que isso não altera substancialmente a freqüência da palavra.

5.2.3 O Corpus ASPA

O ASPA (Avaliação Sonora do Português Atual) está disponível em <http://www.projetoaspa.org/> e corresponde à transcrição fonética de dados do LAEL Escrita. O corpus contabiliza um total de 199.864 tipos e 219.782.537 ocorrências. Em seus dados, foram computados os itens do LAEL Escrita que apresentavam freqüência de ocorrência acima de seis. Durante a elaboração da metodologia de nossa tese, período em que efetivamente consultamos corpora do PB para realizar buscas sobre freqüência de ocorrência, o ASPA ainda estava em fase de transcrição fonética dos dados; em função disso, apesar do seu expressivo número de ocorrências e apesar da vantagem de apresentar os itens foneticamente transcritos, não pudemos contar com esse corpus para selecionar as palavras a serem adotadas em nossos experimentos. Por outro lado, na fase final de redação desta tese, o ASPA já estava em fase de conclusão e buscas preliminares já podiam ser feitas, então, em algumas tabelas desta tese (Cf. Tabelas 5, 8 e 9), utilizamos o ASPA para

contabilizar as freqüências de tipo dos grupos de plural pesquisados, já que esse corpus permite que façamos uma busca por itens terminados em *–ão*, *–l* e ditongo em *–u*. Essa contagem, baseada em dados reais de uso da língua, permitiu uma avaliação mais precisa sobre a quantidade de palavras pluralizadas através de cada uma das classes sob análise nesta tese.

Nos parágrafos subseqüentes, faremos uma comparação mais direta sobre os corpora que permitem busca por freqüência de ocorrência, a fim de justificarmos, de forma empírica, nossa opção pelo Corpus NILC/São Carlos.

5.3 A opção pelo Corpus NILC/São Carlos

Tomada a decisão de selecionar as palavras de nossos experimentos com base na freqüência de ocorrência, outra questão se apresentou: em que corpus do PB iríamos basear nossa contagem de freqüência? Nossas opções para buscas por freqüência de ocorrência, conforme vimos nas subseções anteriores, eram o Corpus NILC/São Carlos e o LAEL, já que o ASPA ainda estava em fase de elaboração quando a metodologia desta tese foi definida.

Um primeiro critério adotado para a escolha de um dos dois corpora foi a quantidade de dados que cada um deles disponibiliza. Em nossa tese, como estamos analisando itens que, em termos de classe, apresentam baixa freqüência de tipo, é importante que o corpus tenha um número significativo de dados, a fim de que as ocorrências de palavras pertencentes aos grupos em *–ão*, *–l* e ditongo em *–u* sejam registradas. Com relação ao quantitativo de dados, o NILC/São Carlos tem 41.372.943 itens, enquanto o LAEL tem 1.182.994. De acordo com esse parâmetro, portanto, o Corpus NILC/São Carlos tem prevalência sobre o LAEL, pois seus dados são 35 vezes em maior quantidade que os do LAEL.

Por outro lado, um segundo critério importante para a escolha do corpus é que ele seja representativo, também, da língua falada, já que é nessa modalidade que as variações lingüísticas se originam. Com relação a esse parâmetro, o Corpus LAEL tem preferência, porque ele disponibiliza dados de fala e de escrita, enquanto o NILC/São

Carlos tem itens apenas de escrita. No entanto, os dados de fala do LAEL contabilizam apenas 16,7% do total das ocorrências do corpus, então não se pode afirmar que esse quantitativo seja altamente expressivo.

Para tentar sanar as dificuldades encontradas na seleção do corpus mais apropriado para nossa análise (LAEL Fala, LAEL Escrita ou Corpus NILC/São Carlos), resolvemos fazer uma análise comparativa. Já havíamos pesquisado, no ASPA e no Novo Dicionário Aurélio Eletrônico, todas as palavras pluralizadas em *–ões*, *–ãos* e *–ães* (Tabelas 5 e 6, no Capítulo 3, "Revisão de Literatura"). Conforme vimos, desses morfemas, *–ãos* e *–ães* apresentavam freqüência de tipo bem inferior à de *–ões*. Por causa disso, pesquisamos, no LAEL Fala, no LAEL Escrita e no Corpus NILC/São Carlos, a freqüência de ocorrência de todas as palavras que se flexionavam em *–ãos* e *–ães* etimológicos. Já que esses morfemas apresentavam baixa quantidade de tipos, essa busca foi possível sem dificuldades. Montamos, então, um banco de dados com a freqüência de ocorrência de todas as palavras em *–ãos* e *–ães* nos corpora mencionados. A partir desses dados, selecionamos as cinco palavras mais freqüentes desse grupo (de plurais em *–ãos* e *–ães*) e montamos uma tabela comparativa sobre os corpora com esses números. Gostaríamos de ter feito a mesma comparação para os demais plurais em análise nesta tese, no entanto, como os demais grupos apresentavam freqüência de tipo maior (Cf. Tabelas 8, 9, 10 e 11), essa busca de freqüência de ocorrência para cada uma das palavras se tornou inviável. Ficamos, então, apenas com a comparação sobre os dados pluralizados em *–ãos* e *–ães*. Apresentamos os resultados na tabela a seguir:

Tabela 17: Comparação entre as cinco palavras mais freqüentes de plural em *–ãos* e *–ães* nos corpora NILC/São Carlos e LAEL (Escrita e Fala)

Corpora	Palavras	Ocorrências
NILC/São Carlos	mãos	2.949
	órgãos	1.578
	irmãos	1.125
	alemães	1.025
	cidadãos	887
LAEL Escrita	mãos	170
	órgãos	37
	cidadãos	22
	irmãos	15
	grãos	09
LAEL Fala	mãos	08
	alemães	05
	órgãos	05
	cidadães	04
	irmãos	04

Na tabela acima, percebemos que, nos três corpora consultados, as cinco palavras mais freqüentes são basicamente as mesmas. As únicas diferenças são: 1) No LAEL Escrita, a palavra "alemães" não figura entre as mais freqüentes. Em quinto lugar nesse corpus está o item "grãos", que não aparece nas demais listas; 2) No LAEL Fala, a palavra "cidadães" (variação de "cidadãos") figura na lista. À parte essas divergências, as cinco palavras mais recorrentes nos três corpora são as mesmas.

Quanto à comparação entre os números de ocorrências das palavras, observamos, nitidamente, que, mesmo que os itens mais freqüentes (em *–ãos* e *–ães*) tenham sido basicamente os mesmos, o quantitativo maior de dados do Corpus NILC/São Carlos se reflete na freqüência de ocorrência das palavras consultadas. Nos três corpora, o item mais freqüente é "mãos", sendo que, no NILC/São Carlos, sua freqüência é de 2.949 ocorrências, no LAEL Escrita, é de 170 ocorrências e, no LAEL Fala, apenas oito. Essa grande diferença numérica é similar em relação a todas as palavras dos três corpora, da primeira à quinta palavra mais freqüente. Essas discrepâncias numéricas refletem a diferença na quantidade de dados dos corpora comparados. O NILC/São Carlos tem 41.372.943 dados, o LAEL Escrita tem 985.093 e o LAEL Fala tem 197.901. Considerando em conjunto o LAEL Escrita e o LAEL Fala, eles apresentam 35 vezes menos dados que o

Corpus NILC/São Carlos, daí as grandes diferenças numéricas quanto às palavras mais freqüentes nos grupos pluralizados em *–ãos* e *–ães*. A partir dessa comparação, percebemos que, como nossa tese analisa grupos de palavras com baixa freqüência de tipo, é necessário adotar um corpus que contenha maior número de dados, a fim de que possamos encontrar, nesse corpus, quantidades significativas de ocorrências para cada uma das classes de plural sob análise. Em um corpus como o LAEL (Fala ou Escrita), como existe um número bem menor de dados que no NILC/São Carlos, seria inviável dividir as palavras em faixas de freqüência baixa, média e alta, porque a maioria das palavras dos grupos investigados já apresenta, tipicamente, baixa freqüência de ocorrência. Pode ser, inclusive, que não encontremos, para os grupos de plurais investigados, um quantitativo suficiente de palavras para preencher cada uma das faixas de freqüência, devido, justamente, à baixa quantidade total de dados nesses dois corpora (LAEL Fala e LAEL Escrita). Um corpus maior, como o NILC/São Carlos, poderá apresentar maior quantidade de dados para as palavras individualmente, possibilitando, assim, a separação dos itens léxicos em diferentes faixas de freqüência. Assim, nossa opção pelo Corpus NILC/São Carlos para a contagem de freqüência de ocorrência em nossa tese ocorreu porque a maior quantidade de dados disponibilizada por esse corpus permite que mais variações nas classes de plural analisadas sejam registradas.

Sanada a questão da opção por um corpora de análise, outro problema relativo à contagem de freqüência se apresentou: para cada palavra, havia duas possibilidades diferentes de freqüência: de singular e de plural. Como nosso objetivo era organizar as palavras em faixas de freqüência, a dificuldade que encontramos foi: qual das duas possibilidades de freqüência (singular ou plural) seria considerada para essa divisão? Após algumas considerações, resolvemos optar pela freqüência de plural, por dois motivos: 1) Conforme o Modelo de Redes postula, as palavras são estocadas como unidades inteiras, então as formas de plural são listadas no léxico do falante, estabelecendo interconexões morfológicas com as formas de singular. Portanto, a freqüência do plural deve ser considerada, já que itens pluralizados estão armazenados no léxico e é sua freqüência (não a dos itens singulares isoladamente) que pode interferir nas generalizações de plural; 2) Se as teorias de representações mentais postulam que palavras mais freqüentes são acessadas mais facilmente, devemos adotar o número de plural, porque o fato de a palavra ser freqüente na sua forma plural pode ajudar na manutenção de sua flexão, evitando possíveis generalizações.

Uma última dificuldade quanto às medidas de freqüência de ocorrência foi encontrada: como definir, para os fenômenos analisados, o que poderiam ser consideradas faixas de freqüência baixa, média e alta? Com relação aos valores numéricos adotados como limiar para cada uma das faixas de freqüência, Bybee (2006) afirma o seguinte:

> The impossibility at the moment of specifying ranges for extreme high, medium and low is only a function of the state of our knowledge. As more empirical studies appear, absolute frequency ranges for each phenomena (*sic*) will eventually be specifiable. (BYBEE; 2006, p. 06)

Então, não há, até o presente momento, definições numéricas precisas para contagens de freqüência, mas precisávamos encontrar alternativas para dividir as palavras em faixas de freqüência, mesmo que o Modelo de Redes não apresentasse as implicações teóricas e práticas para essa definição. Avaliamos, inicialmente, que a divisão das faixas de freqüência deveria levar em consideração a interação entre as freqüências de tipo e de ocorrência dos itens léxicos que compunham as classes de plurais sob análise. Já havíamos feito, anteriormente, buscas de freqüência de tipo para as palavras terminadas em *–ão*, *–l*, e ditongo em *–u*. Essa busca foi realizada em um corpus de uso da língua (ASPA, conforme Tabelas 5, 8 e 9 do Capítulo 2, "Revisão de Literatura"), e no Novo Dicionário Aurélio Eletrônico (Cf. Tabelas 6, 10 e 11, no mesmo capítulo). Com essa pesquisa, já havíamos observado que os três grupos de plurais sob análise nesta tese (*–ão*, *–l*, e ditongo em *–u*) apresentavam baixa freqüência de tipo nos dois corpora consultados, sendo a classe em ditongo em *–u* menos freqüente que as demais (em termos de tipo). Portanto, a busca por tipos já havia sido realizada e era necessário fazer a pesquisa por ocorrências, a fim de utilizar essas duas fontes de dados para a definição das faixas de freqüência. Para essa busca por freqüência de ocorrência, recorremos ao Corpus LAEL (Fala e Escrita), já que ele disponibiliza uma lista de todos os itens léxicos de que é composto, da palavra mais freqüente à menos freqüente, nas modalidades falada e escrita[60]. Apresentamos abaixo os resultados para essa busca:

[60] O Corpus NILC/São Carlos não foi adotado para esta busca porque ele não disponibiliza uma listagem de suas palavras em ordem de freqüência, como o LAEL. Por causa disso, não foi possível definir as palavras mais freqüentes (dos grupos em *–ão*, *–l* e ditongo em *–u*) no Corpus NILC/São Carlos.

Tabela 18: As palavras mais freqüentes de cada grupo de plural, no LAEL Fala e no LAEL Escrita

	LAEL Fala[61]			LAEL Escrita		
Grupos de Plural	**Palavras**	**Freq.**	**% no corpus**	**Palavras**	**Freq.**	**% no corpus**
Plural em *–ão*	condições	41	0,02	informações	573	0,05
Plural em *–l*	hospitais	16	0,008	principais	239	0,02
Plural em dit. *–u*	graus	06	0,003	graus	18	0,001

Através da comparação dos dados da tabela acima, percebemos que os três grupos de plurais apresentam quantitativos numéricos diferenciados nos corpora consultados. No entanto, observamos que, nos dois corpora, a distribuição dos grupos por ordem de freqüência é exatamente a mesma: os itens em ditongo em *–u* são os menos freqüentes, o grupo em *–l* está na faixa intermediária e a classe mais freqüente é a de *–ão*. O diferencial numérico decorre, portanto, da diferente quantidade de dados dos corpora (o LAEL Fala apresenta 197.901 dados e o LAEL Escrita, 985.093). Apesar dessas diferenças numéricas, observamos que, no percentual em relação ao total do corpus, as palavras mais freqüentes de cada classe de plural apresentam quantitativos similares. No LAEL Fala, essa porcentagem varia de 0,003% a 0,02%; no LAEL Escrita, as porcentagens são maiores, entre 0,001% e 0,05%, mas também não existe grande diferencial entre os três grupos de palavras analisadas. Nenhuma das porcentagens das classes de plural ultrapassa, portanto, 0,05% do total dos corpora. Por causa desse percentual similar em relação ao total de dados nos dois corpora consultados e devido à baixa freqüência de tipo dos três grupos de plurais sob análise, optamos por adotar as mesmas faixas de freqüência para todos os plurais. Ficamos, portanto, com as seguintes medidas: freqüência baixa entre zero e 99 ocorrências, freqüência média entre 100 e 500 ocorrências, freqüência alta com mais de 500 ocorrências. Estamos cientes de que os três casos sob análise apresentam diferentes freqüências de tipo e de ocorrência (Cf. tabelas mencionadas acima), no entanto, como o Modelo de Redes não define medidas precisas que determinem os limiares entre as freqüências, optamos por adotar os mesmos valores para todos os grupos. Escalonar diferentes faixas de freqüência a partir da quantidade de tipos e ocorrências para cada uma das classes de plural analisadas demandaria um conhecimento ainda não definido pelo Modelo de Redes, por isso

[61] As porcentagens fornecidas referem-se ao quantitativo de dados dos dois corpora on-line: o LAEL Fala apresenta 197.901 dados e o LAEL Escrita, 985.093. Nessa tabela, foram considerados somente os substantivos e adjetivos flexionados a partir de cada um dos grupos de plurais.

mantivemos os mesmos parâmetros de freqüência de ocorrência para todos os grupos de plurais, mesmo cientes de que essa opção é questionável.

A próxima subseção discute a codificação dos dados e as hipóteses para cada variável considerada na análise.

5.4 Variáveis analisadas para cada grupo de plurais

Fornecidas as informações sobre os experimentos e sobre os corpora do PB consultados, precisamos esclarecer como os dados coletados foram codificados. Assim, organizamos as subseções seguintes, em que apresentamos as variáveis utilizadas para codificar cada um dos grupos de plurais.

5.4.1 Análise das palavras terminadas em *–ão*

Quanto às palavras em *–ão*, algumas variáveis foram selecionados a fim de que pudéssemos encontrar justificativas para a escolha dos falantes pelos plurais em *–ãos*, *–ães* ou *–ões*. Esses fatores foram:

1. Forma de plural aplicada (variável dependente);
2. Número de sílabas;
3. Tonicidade;
4. Estrutura morfológica;
5. Plural etimológico;
6. Segmento precedente;
7. Número de plurais possíveis;
8. Freqüência de ocorrência;
9. Palavra.

Comentaremos cada um desses fatores separadamente.

1. Forma de plural aplicada

Esta é a variável dependente adotada em nossa análise e tem como objetivo apresentar percentuais sobre os morfemas de plurais adotados pelos informantes. Sendo assim, os critérios eram:

a) Plural em *–ões* adotado: "cartão"/"cartões"[62]
b) Plural em *–ãos* adotado: "cristão"/"cristãos"
c) Plural em *–ães* adotado: "cão"/"cães"
d) Ausência de plural: "avião∅"
e) Outra marca de plural: "religião"/"religiõe"

Esta variável representa o tipo de plural que o falante pronunciou nos experimentos, independentemente de essa forma ser considerada "certa" ou "errada" pelas gramáticas normativas do PB. Para elucidarmos melhor esse ponto, vamos utilizar a palavra "escrivão" como exemplo. De acordo com a gramática normativa, seu plural etimológico seria "escrivães", no entanto alguns falantes, em nossos experimentos, adotaram a forma "escrivões". Então, nesse caso, se, no experimento, o falante pronunciou o plural "escrivões", o código para "plural em *–ões* adotado" foi utilizado, mesmo que esse não seja o plural recomendado pelas gramáticas e dicionários do PB. Se não fizéssemos a codificação dessa maneira, não haveria como contabilizar a quantidade de vezes que cada um dos tipos de plural foi adotado pelos informantes consultados.

Codificamos, também, outros dois casos: 1) Situações em que o informante, apesar de estar lendo uma frase em que, pelo contexto, a palavra deveria ser pluralizada, optou por mantê-la em sua forma singular; 2) Situações em que o informante utilizou uma forma de plural diferente dos três morfemas (*–ões*, *–ãos* e *–ães*) e da marca zero. Casos como

[62] Na apresentação da codificação para os dados, exemplificamos apenas uma possibilidade de plural a ser adotada pelos falantes. No entanto, nos experimentos, o informante poderia adotar outras formas de plural, como "cartãos" e "cartães" para "cartão". Nesses casos, a codificação seria em função dessas formas de plural pronunciadas pelo falante. As possibilidades de plural listadas no corpo do texto são somente exemplos com o objetivo de demonstrar como as palavras foram codificadas. Tais exemplos não significam que as formas apresentadas são as únicas possíveis para os plurais em questão.

esse são bastante raros (Cf. SCHERRE, 1988), mas adotamos esse código para captar as peculiaridades dessa flexão de plural.

Quanto à nossa hipótese para essa variável, supúnhamos que, por causa da alta freqüência do morfema *–ões* para pluralizar as palavras terminadas em *–ão* no singular, essa forma (*–ões*) seria a mais comumente adotada pelos falantes.

2. Número de sílabas

Com este critério, investigávamos se o tamanho da palavra teria algum efeito no tipo de plural adotado pelo falante. As opções eram:

a) Palavras monossílabas: "mão", "pão"
b) Palavras com mais de uma sílaba: "caminhão", "exposição", "órgão"

No grupo de plurais em *–ão*, todos os itens monossílabos são pluralizados através dos morfemas *–ãos* e *–ães*. Em função disso, adotamos a variável número de sílabas a fim de verificar se, nas formas de plurais adotadas nos experimentos, os falantes preservariam essa informação sobre a pluralização dos itens monossílabos. Em consonância com o Modelo de Redes e a Teoria dos Exemplares, supomos que as palavras inteiras são estocadas no léxico mental e que generalizações morfológicas emergem a partir de similaridades entre essas palavras. Com base nesses pressupostos, nossa expectativa era de que os falantes preservassem mais os plurais etimológicos em palavras de uma sílaba, já que uma generalização importante sobre esse grupo é que todos os itens são pluralizados através dos morfemas *–ãos* e *–ães*. Portanto, esperávamos que houvesse menos variações de plurais na classe dos monossílabos.

3. Tonicidade

Através da tonicidade, investigávamos se o fato de o plural ocorrer em sílaba átona ou tônica faria com que a palavra fosse mais ou menos suscetível a generalizações. As possibilidades de codificação eram:

a) Palavras oxítonas (com plural na sílaba tônica): "alemão", "mão"[63], "televisão"
b) Palavras paroxítonas (com plural na sílaba átona): "bênção", "órgão"

Quanto a esse fator, é importante lembrar que todas as palavras paroxítonas terminadas em *–ão* formam plural em *–ãos*, como "acórdão"/"acórdãos", "bênção"/"bênçãos"[64]) e "órfão"/"órfãos". Sendo assim, nossa hipótese era de que o plural etimológico em *–ãos* seria preservado nessa classe de palavras paroxítonas, visto que todos os itens desse grupo formam plural em *–ãos*.

4. Estrutura morfológica

Outra informação importante avaliada nos experimentos foi a interferência do sufixo na generalização de plurais. As possibilidades consideradas foram:

a) Sufixo: "cartão", "exposição"
b) Não-sufixo: "avião", "pão"[65]

Como supomos que o léxico mental é organizado considerando, entre outros fatores, conexões morfológicas entre as palavras armazenadas, era importante analisar se haveria alguma correlação possível entre adoção de um tipo específico de plural e estrutura morfológica. A maioria dos sufixos terminados em *–ão* no PB forma plural em *–ões*, conforme vimos no Capítulo 2, "Grupo de Plurais"; então, nossa expectativa era de que os falantes preservariam essa informação quando pluralizassem tais itens.

5. Plural etimológico

O fator plural etimológico era importante a fim de verificarmos a direção que as variações de plural seguiam, ou seja, para percebermos se os plurais infreqüentes (em *–ãos* e *–ães*) estavam sendo alvo de generalizações em direção à forma *–ões*, conforme

---

[63] Os monossílabos, como "mão", "pão", "grão", foram incluídos na classe dos tônicos.
[64] Conforme afirmamos anteriormente, a etimologia de "bênção" é *benedictio*, *onis* (plural em *–ões*), no entanto, por ser paroxítona, a única forma de plural considerada correta pela gramática normativa do PB é "bênçãos".
[65] Os monossílabos foram incluídos na classe não-sufixo.

prevíamos, ou se o contrário (plurais em *–ões* migrando para a classe dos *–ãos* e *–ães*) também ocorreria. É importante esclarecer que o plural etimológico aqui se refere à forma plural latina, portanto havia uma única possibilidade para cada palavra:

a) Plural etimológico em *–ãos*: "irmão"/"irmãos", "vulcão"/"vulcãos"
b) Plural etimológico em *–ães*: "cão"/"cães", "pão"/"pães"
c) Plural etimológico em *–ões*: "avião"/"aviões", "caminhão"/"caminhões"

Nossa hipótese, conforme já mencionamos no Capítulo 2, "Grupos de Plurais", era que, em geral, as palavras pluralizadas em *–ãos* e *–ães* sofreriam generalizações para o grupo de *–ões*, porque esse morfema tem alta freqüência de tipo.

6. Segmento precedente

Com relação a questões fonológicas, analisamos se o segmento que antecedia a terminação *–ão* de singular interferiria em generalizações de plural. Sendo assim, adotamos a seguinte codificação para os segmentos precedentes:

a) Consoante: "cão"/"cães", "pavão"/"pavões"
b) Vogal: "avião"/"aviões"), "leão"/"leões"

Quanto a esse fator, não tínhamos nenhuma hipótese preliminar a respeito da possível interferência do segmento que precedia o morfema de plural.

7. Número de plurais possíveis

Outro fator importante a ser analisado era a quantidade de plurais que a palavra poderia adotar. Como vimos no Capítulo 3, "Revisão de Literatura", algumas palavras que apresentavam plural etimológico em *–ãos* ou *–ães* desenvolveram um plural

analógico em *–ões*, por isso certos itens apresentam, atualmente, dois ou até mesmo três plurais distintos[66]. As possibilidades, portanto, eram:

a) Um plural: "mãos", "pulmões"
b) Dois plurais: "guardiães"/"guardiões"
c) Três plurais: "piãos"/"piães"/"piões", "vulcãos"/"vulcães"/"vulcões"

Nossa hipótese era que, quando a palavra tinha dois ou três plurais possíveis, a forma *–ões* seria a preferida, por causa da sua alta freqüência de tipo. No entanto, poderia ser também que, se a palavra apresentasse plural abundante e a forma etimológica tivesse alta freqüência de ocorrência, não houvesse uma margem muito alta de generalizações, porque supomos que a alta freqüência de ocorrência torna os itens léxicos mais autônomos e imunes a mudanças analógicas.

8. Freqüência de ocorrência

Um dos fatores mais importantes sob análise nesta tese era a interferência da freqüência de ocorrência do item lexical. Conforme mencionamos diversas vezes ao longo do texto, a freqüência é apontada pela literatura lingüística como um dos critérios utilizados pelo falante para categorizar e acessar informações lingüísticas. Por causa disso, dividimos as palavras terminadas em *–ão* em três faixas diferentes de freqüência:

a) Freqüência baixa (de zero a 99 ocorrências): "escorpião", "limão"
b) Freqüência média (de 100 a 500 ocorrências): "leão", "pão"
c) Freqüência alta (mais de 500 ocorrências): "alemão", "órgão"

Nossa expectativa era de que as palavras mais freqüentes conservariam seus plurais etimológicos e as palavras menos freqüentes favoreceriam as variações de plurais. Em nossos experimentos, a faixa de freqüência adotada foi a da palavra no plural. Conforme afirmamos anteriormente, todas as consultas de freqüência para essa codificação

[66] Como vimos na subseção 2.2.2, as gramáticas divergem quanto às palavras que apresentam duas ou três formas diferentes de plural. Por causa disso, consideramos que a palavra possuía dois ou três plurais se pelo menos uma das gramáticas consultadas na subseção 2.2.2 admitisse essa possibilidade.

foram feitas no Corpus NILC/São Carlos, disponível em <www.linguateca.pt/ACDC/>. Maiores detalhes sobre esse corpus já foram mencionados na subseção 5.2.1 deste capítulo.

Ainda quanto à divisão por freqüência, tentamos encaixar palavras com plurais etimológicos diversos em cada uma das faixas, de modo a tornar os dados comparáveis entre si. Algumas vezes, por limitações referentes ao léxico, isso não foi possível.

9. Palavra

Em nossa análise, adotamos um código diferente para cada uma das palavras selecionadas para nossos experimentos. Nosso objetivo era observar se haveria itens léxicos favorecendo cada um dos morfemas de plural para o grupo em *–ão*. Além disso, analisávamos, também, se, tal como a Difusão Lexical (WANG; CHENG, 1977; HSIEH, 1977; OLIVEIRA, 1997) postula, as palavras submetem-se às variações de forma lexicalmente gradual, havendo, portanto, líderes e retardatários nos processos lingüísticos.

Nossa hipótese para esse fator era que as palavras apresentariam comportamentos divergentes no que diz respeito à migração das classes de plurais. Considerávamos, também, que fatores como a freqüência de ocorrência e a familiaridade do item léxico poderiam determinar a suscetibilidade da palavra às generalizações de plural.

Após apresentarmos e comentarmos todas as variáveis relevantes para a análise das palavras terminadas em *–ão*, podemos passar para a próxima subseção, na qual mencionaremos as variáveis adotadas para codificar as palavras terminadas em *–l* e *–u*.

5.4.2 Análise das palavras terminadas em *–l* e ditongo em *–u*

Quanto às palavras terminadas em *–l* e ditongo em *–u*, outras variáveis foram adotadas a fim de investigar se os falantes aplicariam a pluralização em *–l* ou *–u* para as palavras analisadas. Apesar de apresentarmos essa codificação em apenas uma subseção, em termos estatísticos analisamos em separado os itens terminados em *–l* e os terminados em ditongo em *–u*, de modo que pudéssemos observar de forma mais detalhada as peculiaridades de cada uma dessas classes. No entanto, para evitar redundâncias, já que os fatores são basicamente os mesmos para os dois grupos, redigimos apenas uma subseção

sobre as variáveis consideradas para essas duas classes. Minúcias relativas à codificação para esses grupos serão mencionadas quando for necessário.

As variáveis lingüísticas selecionadas foram:

1. Tipo de plural adotado (variável dependente);
2. Estrutura morfológica;
3. Tonicidade;
4. Número de sílabas;
5. Vogal precedente;
6. Freqüência de ocorrência;
7. Palavra.

Cada uma dessas variáveis será explorada abaixo.

1. Tipo de plural adotado

Essa era a variável dependente a ser analisada nesse grupo de palavras. Mais uma vez, convém salientar que esse código referia-se ao plural que o informante adotou, e não ao plural considerado correto pela gramática normativa. Por exemplo: se, para a palavra "troféu", o informante adotou o plural "troféis" (aplicando a regra de pluralização em *–l*, em vez de *–u*), foi utilizado o código para "plural em *–l* adotado". As opções de codificação, portanto, foram:

a) Plural em *–l* adotado: "sal"/"sais"
b) Plural em *–u* adotado: "véu"/"véus"
c) Ausência de plural: "troféuØ"
d) Outra forma de plural adotada: "móvel"/"móvis"

Os itens "c" e "d" acima foram considerados porque, na coleta de dados, houve casos em que, apesar de a frase apresentar figuras claramente no plural, o falante adotou a forma singular; da mesma maneira, houve algumas ocorrências em que os falantes pronunciaram formas não convencionais de plural, como o caso de "móvis" para "móveis" ou de "míssis" para "mísseis".

Conforme já afirmamos anteriormente, nossa expectativa era de que, por causa da alta freqüência de tipo da classe em *–l*, as palavras terminadas em *–u* poderiam estar sujeitas a generalizações em direção ao grupo terminado em *–l*.

2. Estrutura morfológica

A variável estrutura morfológica foi considerada porque queríamos analisar se o fato de a palavra ter sufixo interferiria na preservação ou mudança da regra de plural. Os códigos adotados foram:

a) Sufixo: "europeu", "infantil", "responsável"
b) Não-sufixo: "anel", "céu", "degrau"

No caso das palavras terminadas em ditongo em *–u*, apenas o morfema *–eu* (`[ew]` ou `[εw]`) representa sufixo (origem, como "europeu", "judeu", ou aumentativo, como "fogaréu", "povaréu"), (Cf. subseção 2.3.2, no Capítulo 2, "Grupos de Plurais"); em função disso, apenas a terminação *–eu* (nos casos em que representava sufixo) recebeu esse código. Os monossílabos foram classificados como não-sufixo.

Nossa hipótese aqui era que palavras com sufixo teriam representações mais fortes no léxico mental e, por isso, estariam menos propensas a generalizações de plural, nas duas classes analisadas (em *–l* ou ditongo em *–u*).

3. Tonicidade

Com o critério tonicidade, nosso objetivo era avaliar se haveria diferença na generalização em função da sílaba em que o plural recaía. As possibilidades eram:

a) Palavras oxítonas (com plural na sílaba tônica): "azul", "pastel"
b) Palavras paroxítonas (com plural na sílaba átona): "difícil", "acessível"

Estes códigos eram válidos somente para palavras terminadas em *–l*, uma vez que todas as palavras terminadas em ditongo em *–u* já são, naturalmente, oxítonas, como,

por exemplo, "berimbau", "degrau, "judeu", "troféu". Os monossílabos foram incluídos na classe de plural em sílaba tônica.

Nossa hipótese para esse fator era que palavras que apresentavam plural em sílaba tônica (oxítonas) sofreriam menos generalizações, por conta da prevalência da tonicidade na cadeia sonora da fala.

4. Número de sílabas

Mais uma vez, utilizamos o critério número de sílabas a fim de verificar se o tamanho da palavra causaria alguma interferência nas generalizações de plural. Consideramos as seguintes possibilidades:

a) Palavras monossílabas: "meu", "sol"
b) Palavras com mais de uma sílaba: "azul", "difícil", "europeu, "museu"

A exemplo da codificação adotada para os itens de plural em *–ão*, utilizamos, também para os plurais em *–l* e ditongo em *–u*, apenas duas possibilidades de número de sílabas. Uma divisão em quantidades maiores de sílabas fatiaria muito os dados, deixando um número pequeno de itens para cada fator, já que havia 27 palavras terminadas em *–l* e 19 terminadas em ditongo em *–u*.

Para essa variável, nossa expectativa era que os itens monossílabos fossem menos suscetíveis a generalizações de plural, porque têm um número pequeno de sílabas e, por isso, qualquer mudança é muito saliente.

5. Vogal precedente

Investigamos, também, se o tipo de vogal que antecedia o segmento *–l* ou *–u* no fim da palavra interferiria em possíveis generalizações de plural. As possibilidades consideradas foram:

a) Vogal [a] precedente: "jornal", "mau"
b) Vogal precedente [ɛ]: "chapéu", "mel"
c) Vogal precedente [e]: "meu", "responsável"

d) Vogal precedente [i]: "difícil", "infantil"

e) Vogal precedente [ɔ]: "espanhol", "sol"

f) Vogal precedente [o]: o único item era "gol"

g) Vogal precedente [u]: o único item era "azul"

O grupo das palavras terminadas em *–u* apresenta pequena quantidade de itens (Cf. Tabelas 9 e 11, no Capítulo 3, "Revisão de Literatura"), por isso pudemos contar apenas com palavras terminadas em *–au* e *–eu*. Segundo o Novo Dicionário Aurélio Eletrônico, o grupo em *–iu* tem nove palavras, o grupo em *–ou* tem 12 e o em *–uu* tem sete. Todas essas palavras são praticamente desconhecidas (como, por exemplo, "ararapitiu", "clou", "nambuu") e de freqüência de ocorrência plural zero nos corpora consultados. Por causa disso, para o grupo de plurais em *–u*, resolvemos considerar apenas as palavras terminadas em *–au* e *–eu*, por serem essas mais freqüentes que as demais.

O grupo de plurais terminados em *–l* é mais numeroso (Cf. Tabelas 8 e 10, no Capítulo 3, "Revisão de Literatura"), por isso não tivemos problemas quanto à escolha de itens léxicos para cada um dos segmentos precedentes. Com relação à vogal precedente, não tínhamos nenhuma hipótese preliminar sobre os resultados.

6. Freqüência de ocorrência

Mais uma vez, codificamos as palavras de acordo com sua freqüência de ocorrência no plural, a fim de verificar se palavras mais usadas seriam menos suscetíveis a generalizações de plural. As faixas de freqüência consideradas foram:

a) Freqüência baixa (de zero a 99 ocorrências): "mausoléu", "sol"

b) Freqüência média (de 100 a 500 ocorrências): "céu", "lençol"

c) Freqüência alta (mais de 500 ocorrências): "hospital", "museu"

As faixas de freqüência foram consultadas no Corpus NILC/São Carlos, para o qual já fornecemos detalhes anteriormente. Conforme já mencionamos na codificação para as palavras terminadas em *–ão*, nossa hipótese era que as palavras menos freqüentes,

devido à sua menor acessibilidade no léxico mental, seriam mais suscetíveis a variações de plural.

7. Palavra

Para o grupo de palavras terminadas em *–l* e em ditongo em *–u*, também acrescentamos um código para cada um dos itens adotados em nossos experimentos. Nossa expectativa era que as palavras apresentariam comportamentos diferenciados com relação à migração de plurais, já que, conforme a Difusão Lexical prevê, as mudanças são graduais e atingem o léxico item a item.

Finalizamos, assim, a descrição das variáveis lingüísticas analisadas em nossa tese. A próxima seção apresenta os fatores sociais que foram adotados.

5.5 Fatores sociais

Além de todos os fatores estruturais mencionados acima, consideramos que, em nossa análise, seria imprescindível incluir, também, fatores sociais. Em consonância com Labov (1972), defendemos que o aspecto social é de suma importância em uma pesquisa lingüística, visto que o indivíduo a ser analisado está sócio e historicamente inserido em uma cultura e isso, em maior ou menor escala, interfere em seus comportamentos, inclusive em termos lingüísticos. Por outro lado, se supomos, de acordo com teorias de estocagem do léxico mental, que a gramática emerge do uso que os falantes fazem dela (HOPPER, 1998), e que a experiência do indivíduo afeta as representações mentais (BYBEE, 2001), é possível afirmar que o contexto social no qual o falante convive desempenha algum papel na sua forma peculiar de se colocar no mundo, o que inclui, obviamente, seu comportamento lingüístico. Devido a esses motivos, incluímos variáveis sociais nesta análise. Apresentaremos, abaixo, tais fatores selecionados para esta pesquisa. Devemos mencionar, também, que os mesmos critérios foram utilizados tanto para a codificação dos dados quanto para a escolha dos informantes.

1. Faixa Etária:

- Jovens: entre 15 e 20 anos
- Medianos: entre 35 e 40 anos
- Adultos: entre 55 e 60 anos

A literatura lingüística demonstra que faixa etária é um dos grandes fatores que interferem na fala dos indivíduos. Chambers (1995) observa que, quando existe mudança lingüística, geralmente os falantes mais jovens são os primeiros a adotarem a forma inovadora. Scherre (1998) registra que são os falantes de 7 a 14 anos os que mais apagam as marcas formas de concordância nominal, mas o padrão curvilinear (que diferencia as faixas etárias) não é muito acentuado.

Em nossa pesquisa, selecionamos três faixas etárias a fim de investigar se as generalizações de plural analisadas nesta tese poderiam ser consideradas mudança em progresso. Se esse fosse o caso, esperaríamos que os falantes da primeira faixa etária (15 a 20 anos) apresentassem uma taxa maior de generalizações de plural que os da terceira faixa (entre 55 e 60 anos) (Cf. CHAMBERS, 1995).

2. Escolaridade:

- Fundamental
- Superior

Scherre (1988) observa que os falantes de nível primário de escolaridade são mais propensos ao apagamento das marcas formais de plural do PB. Votre (2003) indica a relevância da variável escolaridade em fenômenos que são estigmatizados socialmente. Além disso, sabemos que, na cultura escolar do Brasil, existe um ensino bastante voltado para idiossincrasias lingüísticas, como, por exemplo, mostrar aos alunos todas as possibilidades de plurais para as palavras em *–ão* e incentivá-los a memorizar as irregularidades. Por causa disso, resolvemos adotar o fator escolaridade a fim de verificar se falantes que freqüentaram a escola por mais tempo tenderiam a preservar mais as formas padronizadas de plural do que indivíduos que têm nível mais baixo de instrução.

Com relação à interação idade/escolaridade, para os falantes de 15 a 20 anos, tivemos de contar com aqueles que haviam, apenas, ingressado no Ensino Superior, já que

aos 20 anos de idade é difícil que alguém já tenha concluído essa etapa escolar. Nas demais faixas etárias, fomos bastante criteriosos quanto ao nível de escolaridade dos informantes, considerando apenas aqueles que, efetivamente, já haviam concluído o Ensino Fundamental ou Superior.

Nossa hipótese para esse fator, em consonância com a literatura lingüística, era que falantes mais escolarizados teriam menor probabilidade de generalizações de plural. No entanto, a questão da estigmatização dos fenômenos lingüísticos pode ser uma variável importante nesse processo de migração de plurais. Não fizemos teste de reação subjetiva, a fim de verificar se a utilização de um plural por outro (nos grupos de palavras analisados) teria algum juízo de valor. Mesmo assim, esperávamos que falantes mais escolarizados preservassem mais os plurais etimológicos, por causa do esforço escolar em manter a norma padrão.

3. Gênero:

- Feminino
- Masculino

Na literatura sociolingüística, várias são as análises que apontam comportamentos divergentes entre homens e mulheres no que diz respeito ao uso da linguagem. Labov (1972, p. 243) afirma que "In careful speech, women use fewer stigmatized forms than men, (LABOV 1966a:288), and are more sensitive to the prestige pattern." Chambers (1995) observa que as mulheres usam menos variantes não-padrão, demonstram ter maior consciência do prestígio que a fala possui e modelam melhor a linguagem de acordo com a situação de comunicação. Labov (2001) apresenta um capítulo inteiro sobre o "paradoxo do gênero", que consiste no fato de que as mulheres são mais atentas ao padrão lingüístico, mas costumam, ao mesmo tempo, ser mais progressivas e, conseqüentemente, líderes nas mudanças lingüísticas. Scherre (1998) observa que as mulheres preservam mais as marcas de concordância nominal que os homens. Paiva (2003, p. 36) aponta que

> "Quando se trata de implementar na língua uma forma socialmente prestigiada [...], as mulheres tendem a assumir a liderança da mudança. Ao contrário, quando se trata de implementar uma forma socialmente desprestigiada, as mulheres assumem uma atitude conservadora e os homens tomam a liderança do processo."

Já que há tantas pesquisas indicando a interferência do gênero na linguagem, incluímos esse fator em nossa pesquisa, a fim de investigar se homens e mulheres apresentariam comportamentos lingüísticos divergentes no que concerne às marcações de plural. Nossa expectativa, em consonância com a literatura lingüística, é de que as mulheres tenderiam a preservar mais que os homens as formas padronizadas de plural.

Após esclarecermos todos os fatores sociais que foram considerados em nossa análise, passaremos à próxima subseção, que descreve como o trabalho de campo foi conduzido e aponta algumas peculiaridades referentes à codificação dos dados.

5.6 A realização do trabalho de campo

A coleta de dados ocorreu entre janeiro e março de 2006, na cidade de Nova Friburgo, no estado do Rio de Janeiro[67]. Os informantes que se enquadravam no perfil social delineado foram encontrados através de amigos. Contatamos tais pessoas e as convidamos para participarem do experimento, sem explicar exatamente qual seria o objeto de estudo nesta investigação. Dissemos aos falantes que estávamos fazendo uma pesquisa para a faculdade e precisávamos de pessoas naquela faixa etária, gênero e nível de escolaridade. A justificativa que adotamos para o uso dos experimentos foi que estávamos fazendo um teste de memória e queríamos analisar se a idade interferiria na capacidade de as pessoas se lembrarem de figuras ou palavras. No grupo de informantes consultados, não havia pessoas que eram amigos da entrevistadora, portanto ninguém ficou sabendo exatamente qual era o objeto real de estudo da pesquisa.

Na tabela abaixo, apresentamos a distribuição dos 36 informantes de acordo com os fatores sociais selecionados em nossa análise. A identidade das pessoas foi preservada e utilizamos números correspondentes aos indivíduos entrevistados. Em cada

[67] A opção por essa cidade não tem nenhum significado especial. Ela foi escolhida simplesmente por ser o local onde a pesquisadora residia na época de realização da pesquisa.

grupo de plural, o número de informantes e os códigos utilizados para cada um deles são os mesmos.

Tabela 19: Distribuição dos informantes de acordo com fatores sociais

Jovens – 15 a 20 anos			
Ensino Fundamental		**Ensino Superior**	
Gên. Masculino	**Gên. Feminino**	**Gên. Masculino**	**Gên. Feminino**
20	25	15	35
5	34	23	4
21	16	12	1
Medianos – 35 a 40 anos			
Ensino Fundamental		**Ensino Superior**	
Gên. Masculino	**Gên. Feminino**	**Gên. Masculino**	**Gên. Feminino**
18	30	24	11
28	26	10	33
8	17	19	27
Adultos – 55 a 60 anos			
Ensino Fundamental		**Ensino Superior**	
Gên. Masculino	**Gên. Feminino**	**Gên. Masculino**	**Gên. Feminino**
36	31	29	22
3	2	14	32
7	6	9	13

Através da tabela acima, observamos que os 36 informantes foram homogeneamente divididos de acordo com faixa etária, grau de escolaridade e gênero. Houve exatamente o mesmo número de informantes por célula (três), a fim de que compuséssemos uma amostra comparativamente adequada sobre os grupos sociais analisados.

Com relação à realização das entrevistas, após o contato inicial, a pesquisadora deslocou-se até a casa ou local de trabalho do informante (segundo preferência dele), no horário previamente marcado. Em geral, os informantes foram bastante receptivos com a pesquisadora e colaborativos com a realização dos experimentos. Antes do início dos testes, todos os falantes leram e assinaram o "Termo de Consentimento Livre e Esclarecido" (conforme mencionamos anteriormente) e foram informados de que seria necessário gravar e transcrever os dados da entrevista, mas informações específicas relacionadas à sua identidade pessoal seriam preservadas.

Com esses esclarecimentos feitos, explicamos detalhadamente ao informante como seria a realização dos experimentos e como eles deveriam proceder. Após certificarmo-nos de que o informante havia entendido perfeitamente o que teria de fazer, começamos a gravação dos testes. Cada experimento durou em torno de 10 a 15 minutos, dependendo da agilidade do entrevistando em realizar as tarefas determinadas. A maioria dos falantes ficou visivelmente nervosa ao saber que sua fala estava sendo gravada. Essa apreensão foi mais marcante no momento inicial da interação, mas foi amenizada com o desenrolar do experimento. Em quase todos os informantes vimos muito claramente uma preocupação e/ou insegurança em "acertar" todas as palavras que eram representadas nas figuras. Ao final das 36 entrevistas, contamos com o seguinte quantitativo de dados:

- Palavras terminadas em *–ão*: 1.341 dados;
- Palavras terminadas em *–l*: 1.167 dados;
- Palavras terminadas em ditongo em *–u*: 971 dados.

Em cada uma das etapas da coleta dos dados, houve fatos que demonstraram idiossincrasias dos falantes e nos fizeram, também, refletir sobre a natureza do fenômeno analisado. Comentaremos tais peculiaridades a seguir, porque julgamos que são importantes para nossa pesquisa.

Na primeira parte da interação ("Conjunto de Figuras"), houve comportamentos díspares dos falantes, ao serem questionados sobre o que viam em cada uma daquelas gravuras. Alguns falantes foram extremamente sucintos, dizendo apenas "duas <u>mãos</u>", "três <u>caminhões</u>", e assim sucessivamente. Outros foram mais prolixos e falaram frases inteiras sobre o que estavam vendo: "Na primeira figura, eu vejo duas <u>mãos</u>, uma masculina e uma feminina, a masculina dando suporte à feminina". Houve casos, também, de não-marcação do plural, decorrentes de estruturas como: "um conjunto de <u>berimbauØ</u>", "um monte de <u>grãoØ</u>". Da mesma forma, ocorreram, também, casos em que o falante via as figuras com dois piões ou vários jornais e dizia apenas "piãoØ" ou "jornalØ". Essa característica de usar o singular para referir-se a uma classe como um todo parece ser uma tendência marcante no PB, por isso retomaremos esse ponto no Capítulo 6, "Análise de Dados". Ainda nessa etapa do "Conjunto de Figuras", observamos casos de formas que não se enquadravam ao padrão de pluralização esperado para aquela palavra. Exemplos disso são ocorrências de "móvis" como plural de "móveis" e "míssis" para

"mísseis". Ocorrências semelhantes aconteceram, também, nas demais modalidades de elicitação de dados. Esses casos foram codificados como "Outra forma de plural adotada" e serão, também, discutidos no Capítulo 6, "Análise de Dados".

Inúmeros foram os momentos em que os falantes claramente se sentiam em dúvida quanto ao plural adequado. Palavras como "chapéu", "pião", "sol" geraram grandes questionamentos por parte dos falantes, como "Chapéus ou chapéis? Chapéu, chapéis... não sei qual que é a certa". Nessas situações, ficava claro que havia formas em competição no léxico mental do falante e este se sentia confuso quanto à utilização de um ou outro plural. Casos como esses foram comuns em todas as etapas dos experimentos. Na codificação dos dados, consideramos a resposta final que os falantes adotaram para a palavra, após o momento de hesitação.

Na "Leitura de Frases", os falantes, em geral, não apresentaram dificuldades. As letras eram grandes, de modo a facilitar a leitura pelos informantes que eventualmente sofressem de algum problema de visão. Mais uma vez, as dúvidas dos informantes quanto ao plural de certos itens aconteceram. Casos de não-marcação do plural também foram registrados, como "Será que existem dois solØ na nossa galáxia?" ou "Crianças geralmente gostam de rodar piãoØ.".

O "Teste de Reação" foi a etapa da coleta de dados que gerou maiores dúvidas nos falantes. Nesse experimento, a diversidade e o número de palavras foram maiores, porque podíamos prescindir da representação gráfica dos itens. Sendo assim, havia palavras bastante raras, como "jirau" e "mausoléu", palavras que quase nunca são usadas no plural, como "mel" e "sol", bem como palavras que apresentam formas concorrentes de plural, como "anão" ("anões"/"anães"/"anãos") e "vulcão" ("vulcões"/"vulcães"/"vulcãos"). Por causa dessas características, a dificuldade do teste foi um pouco maior. Pudemos perceber que, para algumas palavras, os falantes formavam o plural prontamente, como "europeu"/"europeus", "pão"/"pães", entre outros. No entanto, principalmente nos casos de palavras pouco freqüentes ou nos casos em que havia mais de uma possibilidade de plural, os falantes hesitaram, demorando mais tempo para falar o plural, oscilando entre duas formas diferentes ("Sols ou sóis?") ou até mesmo alongando o segmento precedente, como "Jirau – jiraaaaaaaus", a fim de ter mais tempo para escolher uma forma de desinência plural. Outro detalhe que observamos foi que algumas formas de plural, como "vulcões" para "vulcão" e "anões" para "anão", já são tão cristalizadas (a ocorrência das formas etimológicas – "vulcãos" e "anãos", respectivamente – é inexistente, segundo o

Corpus NILC/São Carlos), que os falantes não demoraram para formar seus plurais, dada a lacuna dessas formas etimológicas no léxico diário. De forma análoga, para alguns falantes a forma de plural não-padrão já é a cristalizada, por isso "chapéis" e "degrais" foram prontamente apontadas como o plural de "chapéu" e "degrau", respectivamente. Ainda com relação à opção por uma ou outra forma de plural, observamos que alguns falantes não foram consistentes em todas as etapas dos experimentos. A palavra "chapéu", por exemplo, estava presente em todas as modalidades dos experimentos, e houve situações em que um mesmo falante adotou as formas "chapéis" e "chapéus" em fases distintas dos experimentos, sem se dar conta de que estava utilizando plurais diferentes para o mesmo singular. Tais fatos parecem demonstrar a existência de exemplares em competição no léxico mental dos falantes.

Um último comentário a ser feito é quanto à familiaridade da palavra. Observamos que algumas palavras, embora freqüentes, não são familiares, como "míssil"; por causa disso, a formação de plural para esse tipo de palavra apresentou um pouco de dificuldade para os falantes. Uma questão que gostaríamos de mencionar é que a contagem da freqüência de ocorrência da palavra é um excelente recurso para análise em lingüística de corpora; no entanto, em geral os corpora disponíveis para esse tipo de pesquisa baseiam-se em textos escritos ou em interações orais formais, como transcrição de aulas, por exemplo. Por isso, a alta freqüência de um item em um corpus de língua pode não corresponder à alta freqüência no léxico de um falante específico. Além das diferenças de formalidade entre fala e escrita que se refletem em discrepâncias entre um corpus de língua e o léxico do falante, não podemos negar que cada indivíduo tem uma história de vida e suas experiências enquanto ser humano se refletem, também, em seu vocabulário. Nesse sentido, não existe compatibilidade total entre um corpus lingüístico e o léxico de um falante. Toda comparação entre essas duas instâncias limita-se, apenas, a índices de freqüência, não a contagens categóricas.

Os dados coletados nas entrevistas foram digitalizados e editados segundo os parâmetros descritos anteriormente. Organizamos uma matriz no programa Microsoft Excel, a fim de facilitar a codificação dos dados. Para a análise estatística e probabilística, utilizamos o programa SPSS, delineado especificamente para pesquisas de caráter lingüístico. Nesse *software*, o recurso adotado foi a regressão logística, que analisa uma variável dependente binária (com dois fatores, portanto) e tenta identificar as variáveis independentes (mencionadas anteriormente na codificação dos dados) que interferem no

fenômeno em questão. Essa opção pela regressão binária acarretou outras decisões metodológicas, que serão oportunamente esclarecidas, no Capítulo 6, "Análise de Dados".

Após a descrição detalhada da metodologia e das variáveis lingüísticas e extralingüísticas a serem adotadas nesta tese, passaremos ao capítulo seguinte, que analisa os dados coletados para nossa pesquisa, correlacionando-os com as hipóteses levantadas para cada uma das variáveis mencionadas neste capítulo de "Metodologia" e procurando alinhar os resultados com o Modelo de Redes e a Teoria dos Exemplares, escopos teóricos adotados em nossa tese.

CAPÍTULO 6

Análise de dados

Utilizando o Modelo de Redes (BYBEE, 1985, 1995, 2001) e a Teoria dos Exemplares (PIERREHUMBERT, 2000, 2001a, 2001b, 2001c), discutidos no Capítulo 4, "Quadro Teórico", o presente capítulo tem como objetivo apresentar os resultados alcançados com a coleta de dados que realizamos. Apesar de essas duas correntes teóricas serem a linha básica de argumentação seguida ao longo desta tese, outras teorias sobre léxico mental (Modelo Conexionista[68] e Teoria de Palavras e Regras[69]) podem ser eventualmente mencionadas, caso isso se faça necessário a partir da configuração dos dados.

Este capítulo está organizado da seguinte maneira: nas subseções 6.1, 6.2 e 6.3, apresentamos os resultados para cada um dos grupos de plurais analisados (em *–ão*, *–l* e ditongo em *–u*, nessa ordem); na subseção 6.4, comentamos os resultados do "Teste de Reação"; na subseção 6.5, propomos uma discussão geral dos principais resultados alcançados com os experimentos; finalmente, na subseção 6.6, encerramos o capítulo com as conclusões mais relevantes que obtivemos com a análise dos dados.

6.1 Resultados para os plurais terminados em *–ão*

Com relação ao grupo de palavras que termina em *–ão* no singular e se pluraliza através das desinências *–ãos*, *–ães* ou *–ões*, coletamos um total de 1.341 dados através dos experimentos realizados. Apresentamos, abaixo, os resultados alcançados pelo programa SPSS. Elaboramos tabelas e/ou gráficos ilustrativos para cada um desses resultados.

---

[68] Cf. RUMELHART; McCLELLAND, 1986; PLUNKETT; MARCHMAN, 1991; MacWHINNEY; LEINBACH, 1991, ELMAN, 1998b; HARE; ELMAN, 1995; DAUGHERTY; SEIDENBERG, 1994.
[69] Cf. PINKER; PRINCE, 1988, 1994; PINKER, 1991, 1999; PINKER; ULLMAN, 2002a, 2002b.

A variável dependente que consideramos em nossa análise foi o tipo de plural adotado pelo falante. As possibilidades eram: 1) Plural em *–ões*; 2) Plural em *–ãos*; 3) Plural em *–ães*; 4) Ausência de plural[70]. Os resultados gerais para a variável dependente foram:

Tabela 20: Resultados gerais para os plurais terminados em *–ão*

Formas de Plural	**N**	**%**
Plural em *–ões*	805/1.341	60
Plural em *–ãos*	260/1.341	19,3
Plural em *–ães*	214/1.341	15,9
Ausência de plural	62/1.341	4,6

Os resultados gerais indicam preferência dos falantes pelo morfema *–ões* para pluralizar palavras terminadas em *–ão* no singular (60%). As formas *–ãos* e *–ães* foram utilizadas em porcentagens bem menores que *–ões* (19,3% e 15,9% para *–ãos* e *–ães*, respectivamente).

Os 62 casos de ausência de plural (4,6%) ocorreram nos seguintes contextos: 1) No "Conjunto de Figuras", o falante via uma gravura, como, por exemplo, a que continha quatro balões, e dizia "balão", mesmo que a figura indicasse, claramente, que havia mais de um item da mesma classe; 2) Ainda no "Conjunto de Figuras", o falante via uma gravura com itens em número maior que um e utilizava expressões quantificadoras com o substantivo no singular, como "uma frota de caminhão", "vários tipos de grão"; 3) Na "Leitura de Frases", também aconteceu o uso do singular pelo plural, como, por exemplo, "Crianças geralmente gostam de rodar pião", ou "Dirigir caminhão é muito perigoso nas estradas brasileiras".

Conforme afirmamos no Capítulo 2, "Grupos de Plurais", existe, no PB, possibilidade de utilizar a forma de singular indicando toda uma classe. Müller (2003, p. 83) cita exemplos de frases como "Jorge sempre lê revista depois do jantar.", em que uma expressão no singular ("revista") é utilizada para referir-se a um número plural, já que Jorge pode ler diferentes revistas a cada dia, não necessariamente uma mesma revista todos os dias. Thomas (1969) também comenta a utilização do singular pelo plural como uma característica do PB (Cf. subseção 3.1 do Capítulo 3, "Revisão de Literatura"). Parece, então, que os informantes, em alguns casos (4,6%), demonstraram essa tendência à não-

[70] A variável "Outro plural", mencionada no Capítulo 5, "Metodologia", não ocorreu no grupo de plurais em *–ão*, por isso não foi comentada na presente análise.

-marcação do plural, já que existe, no PB, a possibilidade de utilizar o singular genericamente, denotando toda uma classe.

Utilizando os recursos estatísticos disponibilizados pelo programa SPSS, calculamos os números, as porcentagens e os pesos relativos para as variáveis consideradas relevantes na análise dos plurais em *–ão*. O tipo de análise estatística que utilizamos foi a regressão binária, na qual se escolhe uma variável dependente (com apenas duas variantes, já que a regressão é binária) como objeto da investigação e contrapõem-se tais fatores a outras variáveis independentes. Conforme vimos, a variável plurais em *–ão* no singular apresentava quatro possibilidades distintas: 1) Plural em *–ões*; 2) Plural em *–ãos*; 3) Plural em *–ães*; 4) Ausência de plural. Como a regressão binária processa apenas dois fatores, tivemos de recorrer a duas alternativas: exclusão e fusão de códigos. Em primeiro lugar, excluímos os dados de ausência de plural porque os 62 casos de plural não-marcado formalmente representam uma parcela pequena dos dados (4,6%) e porque tais ocorrências já foram exploradas nos parágrafos anteriores. Em segundo lugar, fundimos os plurais de *–ãos* e *–ães* em uma só variável, em função de dois motivos: 1) Esses dados tiveram porcentagens semelhantes de pluralização (19,3% e 15,9%, respectivamente, contra 60% de plurais em *–ões*); 2) A pluralização canônica do PB é feita através do morfema *–s*, no entanto, no grupo de palavras em *–ão*, o plural mais produtivo (que atrai novos membros desde o português arcaico) é *–ões*, em oposição a *–ãos* e *–ães*, que, devido à sua baixa freqüência de tipo, só se mantêm em palavras de alta freqüência de ocorrência (Cf. CAMPOS, 1981). Pode-se argumentar que o plural em *–ãos*, por ser feito a partir da adição de *–s*, é regular, e os plurais em *–ães* e *–ões*, por sofrerem alterações no radical, são irregulares. Então, em função dessas similaridades, talvez a oposição mais coerente fosse *–ãos* x *–ães* e *–ões*. No entanto, como vimos no Capítulo 4, "Quadro Teórico", para o Modelo de Redes, o que define a estocagem das palavras no léxico mental é sua freqüência de ocorrência, não sua estrutura interna (regular ou irregular). Um modelo de regras simbólicas, por sua vez, postularia que palavras irregulares são memorizadas e regulares são derivadas por regras abstratas, portanto, para uma teoria dessa natureza, faria sentido a distinção entre plural regular (*–ãos*) versus plurais irregulares (*–ães* e *–ões*). No entanto, como as correntes teóricas que adotamos em nossa tese (Modelo de Redes e Teoria dos Exemplares) conferem ao uso da língua papel determinante na organização do léxico mental, devemos ser coerentes com esses modelos e fazer a oposição entre o plural mais produtivo (em *–ões*) e os menos produtivos (em *–ãos* e *–ães*), a despeito de sua (ir)regularidade. Bybee (1994) define

produtividade como a capacidade que uma classe paradigmática tem de atrair novos itens. Afirmamos que o plural em *–ões* é mais produtivo e os plurais em *–ãos* e *–ães* são menos produtivos a partir das informações apresentadas nos Capítulos 2 e 3. Conforme vimos, a classe em *–ões* atrai mais membros desde o português arcaico e é adotada nos plurais analógicos. Por outro lado, *–ãos* e *–ães* só sobrevivem em formas que, individualmente, apresentam alta freqüência de ocorrência ou em itens marcados, como monossílabos.

Além dos motivos supracitados, os plurais em *–ãos* e *–ães* representam uma parcela pequena de itens no léxico do português, em oposição a *–ões*, que se aplica a uma quantidade maior de palavras. Na Tabela 5 do Capítulo 3 ("Revisão de Literatura"), já apresentamos a freqüência no ASPA das palavras terminadas em *–ões*, *–ãos* e *–ães*. Naquela tabela, verificamos que *–ões* é aplicado a 1.352 palavras (97,3% do total de palavras do grupo em *–ão* no plural), *–ãos* pluraliza 25 itens (1,7% do total) e *–ães*, 12 palavras (0,8%)[71]. Existe, portanto, um morfema produtivo e freqüente (*–ões*) e dois menos produtivos e infreqüentes (*–ãos* e *–ães*). Em função dessas similaridades decorrentes do uso das formas plurais, fundimos as terminações *–ãos* e *–ães* e as opusemos ao morfema *–ões*. Outra justificativa para esse ajuste nos dados é que a hipótese investigada nesta pesquisa (Cf. Capítulo 4, "Quadro Teórico") é a migração de plurais em *–ãos* e *–ães* para *–ões*, ou seja, estamos analisando, de um lado, *–ãos*/*–ães* contra *–ões*, do outro. Devido a esses motivos, consideramos que tais adequações na recodificação dos dados seriam pertinentes e estariam de acordo com a hipótese de trabalho que investigamos.

Ficamos, então, com um total de 1.279 dados em que uma forma de plural foi, efetivamente, adotada pelo falante. Excluindo os casos de ausência de plural e fundindo as variáveis *–ãos* e *–ães* em uma só, o SPSS selecionou como relevantes as seguintes variáveis: **palavra**, **freqüência de ocorrência**, **faixa etária** e **escolaridade**. Como nossa pesquisa investiga, principalmente, o papel da freqüência nas representações mentais, esse será o primeiro resultado a ser analisado. Com relação a essa variável, os valores de números, porcentagens e pesos relativos[72] são:

---

[71] Na Tabela 6, observamos que o Novo Dicionário Aurélio Eletrônico também ratifica a hipótese de que, para o singular em *–ão*, o morfema mais freqüente é *–ões*.

[72] Os resultados para número (N), porcentagem (%) e peso relativo (P.R.) foram calculados em relação aos plurais em *–ões*, porque, na regressão binária feita pelo programa SPSS, era necessário escolher uma variante (da variável dependente) a partir da qual os pesos relativos seriam apresentados. Como nossa hipótese é de que os plurais em *–ãos* e *–ães* etimológicos migram para *–ões*, decidimos que *–ões* seria a variante a partir da qual os pesos relativos seriam medidos. Assim, nossas tabelas apresentam quantitativos numéricos relativos à aplicação de plural em *–ões*.

Tabela 21: Efeito da freqüência de ocorrência nos plurais em *–ão*

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	344/425	80,9	0
Freqüência Média	223/415	53,7	0.99
Freqüência Alta	238/439	54,2	1

Nos dados acima, os resultados para freqüência de ocorrência são contrários à nossa hipótese inicial de trabalho. De acordo com o Modelo de Redes, esperar-se-ia que as palavras de baixa freqüência, por serem menos disponíveis no léxico mental, apresentassem maior tendência à adoção do plural em *–ões*, tipo mais freqüente para palavras com singular em *–ão*. Contrariamente, em termos de peso relativo, os resultados demonstram que as palavras mais freqüentes adotaram plurais em *–ões*, e as menos freqüentes desfavoreceram esses plurais. Por outro lado, os pesos relativos encontrados são categóricos, variando entre zero para freqüência baixa, 0.99 para freqüência média e 1 para freqüência alta. As porcentagens também não são compatíveis com os pesos relativos, já que, para 80,9%, o peso relativo é zero e, para 54,2%, é 1. Devemos investigar por que esses problemas estatísticos ocorreram e por que os resultados, em termos de pesos relativos, são contrários à nossa hipótese preliminar.

Iniciando nossa investigação, analisaremos as palavras individualmente, para verificar se existem itens léxicos interferindo nos resultados para freqüência de ocorrência. Montamos, então, uma tabela com todas as palavras terminadas em *–ão* utilizadas em nossos experimentos, a faixa de freqüência de cada uma delas (Cf.Corpus NILC/São Carlos), sua etimologia e quantas delas foram pluralizadas com o morfema *–ões* em nossos experimentos[73]:

[73] Nesta tese, todas as tabelas sobre freqüência de ocorrência de palavras apresentam a mesma organização: os dados estão listados em ordem crescente por faixa de freqüência de ocorrência e uma linha horizontal divide as diferentes faixas de freqüência de ocorrência (baixa, média e alta, sempre nessa ordem).

Tabela 22: Palavras adotadas nos experimentos sobre plurais em *–ão*

Palavras	Freqüência	Plural Etimológico	N	%	P.R.
1. Pião	baixa –1	*–ões*	59/61	96,7	0.43
2. Pavão	baixa – 5	*–ões*	34/35	97,1	0.46
3. Limão	baixa – 13	*–ões*	56/61	91,8	0.22
4. Escrivão	baixa – 14	*–ães*	17/35	48,5	0
5. Escorpião	**baixa – 23**	***–ões***	**25/25**	**100**	**1**
6. Bênção	baixa – 26	*–ões*	6/35	17,1	0
7. Guardião	baixa – 26	*–ães*	15/35	42,8	0
8. União	baixa – 34	*–ões*	31/35	88,5	0.16
9. Vulcão	baixa – 62	*–ãos*	101/102	99	0.72
10. Balão	**média – 110**	***–ões***	**34/34**	**100**	**1**
11. Pão	média – 118	*–ães*	0/106	0	0
12. Televisão	**média – 122**	***–ões***	**27/27**	**100**	**1**
13. Leão	**média – 125**	***–ões***	**33/33**	**100**	**1**
14. Anão	média – 142	*–ãos*	24/36	66,6	0
15. Pulmão	**média – 164**	***–ões***	**22/22**	**100**	**1**
16. Religião	média – 191	*–ões*	35/36	97,2	0.47
17. Cristão	média – 235	*–ãos*	12/35	34,2	0
18. Cão	média – 350	*–ães*	0/35	0	0
19. Grão	média – 471	*–ãos*	0/14	0	0
20. Exposição	média – 494	*–ões*	35/36	97,2	0.47
21. Opinião	**alta – 571**	***–ões***	**36/36**	**100**	**1**
22. Caminhão	**alta – 792**	***–ões***	**63/63**	**100**	**1**
23. Avião	**alta – 874**	***–ões***	**104/104**	**100**	**1**
24. Cartão	**alta – 941**	***–ões***	**27/27**	**100**	**1**
25. Alemão	alta – 1.025	*–ães*	5/36	13,8	0
26. Irmão	alta – 1.125	*–ãos*	0/35	0	0
27. Órgão	alta – 1.578	*–ãos*	4/36	11,1	0
28. Mão	alta – 2.494	*–ãos*	0/104	0	0

Na tabela acima, observamos que, das 28 palavras adotadas nos experimentos, houve nove casos de uso categórico de *–ões*. Essas palavras estão em negrito e são: "escorpião", "balão", "televisão", "leão", "pulmão", "opinião", "caminhão", "avião" e "cartão". Como característica comum a esses itens, apontamos o fato de que todos eles já têm plural etimológico em *–ões* e, por isso, apresentam números categóricos de aplicação do morfema *–ões*.

Outra questão a ser notada quanto à tabela acima é o fato de que os pesos relativos não apresentam gradualidade compatível com a distribuição percentual dos dados. Todos os resultados abaixo de 66,6% apresentam peso relativo zero, o que faz com que distribuições percentuais de, por exemplo, 11,1% e 48,5%, apresentem ambas peso relativo zero. Esse resultado não parece ser coerente com a diferença que os números percentuais apresentam.

Devemos relembrar que, no SPSS, a variante a partir da qual os pesos relativos são calculados é plural em *–ões*. Se algumas palavras já têm plurais etimológicos em *–ões*, é natural que apresentem números categóricos de aplicação do morfema *–ões*. Por causa disso, não faz sentido incluir, em um mesmo grupo, palavras em *–ãos* e *–ães*, que podem, potencialmente, adotar plurais em *–ões*, e palavras que já se pluralizam em *–ões* e, portanto, não poderiam migrar para aquilo que, etimologicamente, já são. Como os cálculos de peso relativo são feitos com base no plural em *–ões*, parece que a inclusão de plurais etimológicos em *–ões* está mascarando os dados e impedindo que resultados mais graduais para os pesos relativos sejam alcançados. A fim de obter cálculos mais adequados para esses plurais, seria necessário, então, excluir de nossa análise os itens de etimologia em *–ões*. Essa decisão é discutível e, por isso, antes de concluir definitivamente que é necessário remover esses plurais etimológicos (em *–ões*), é preciso testar se nossa hipótese (de que plurais em *–ãos* e *–ães* estão migrando para *–ões*) se fundamenta através dos dados coletados. Caso isso seja verificado, poderemos eliminar os dados etimológicos em *–ões*, contribuindo, assim, para maior clareza em nossa análise. Vamos observar, então, a distribuição dos dados de acordo com plurais etimológicos. A tabela abaixo apresenta todas as variações que os dados sofreram, de acordo com seus plurais etimológicos:

Tabela 23: Efeito do plural etimológico nos plurais em *–ão*

Plurais Etimológicos	**Respostas em *–ões***		**Respostas em *–ãos***		**Respostas em *–ães***	
	N	**%**	**N**	**%**	**N**	**%**
Etim. *–ões*	627/670	93,5	41/670	6,1	02/670	0,2
Etim. *–ãos*	**141/362**	**38,9**	206/362	56,9	15/362	4,1
Etim. *–ães*	**37/247**	**14,9**	13/247	5,2	197/247	79,7

Os dados da tabela acima são bastante significativos. Observamos que a grande corrente de migração de plurais é, primeiramente, de *–ãos* para *–ões* (38,9%, em negrito na tabela) e, em segundo lugar, de *–ães* para *–ões* (14,9%, também em negrito). Em cada uma

das classes de plurais predominam os plurais etimológicos, ou seja, os plurais em *–ões* foram preservados em 93,5% das respostas, os em *–ãos* foram mantidos em 56,9% e os em *–ães* prevaleceram em 79,7% dos dados. Os plurais que mais conservaram sua etimologia são, portanto, em *–ões*, possivelmente devido à sua alta freqüência de tipo. Os plurais em *–ães* também tenderam a ser preservados, talvez porque haja uma pequena quantidade deles no PB (12, segundo o ASPA, ou seja, 0,8% de todos os dados de *–ão* no plural), o que os torna bastante marcados, e também porque nessa classe há itens monossílabos de alta freqüência de ocorrência ("cães", "pães", por exemplo), que, conforme veremos posteriormente, tendem a manter os plurais etimológicos.

Houve 43 casos de palavras etimológicas em *–ões* que adotaram plurais em *–ãos* e *–ães*, o que corresponde a 6,3% do total de dados para *–ões*. Tais casos foram: "bênçãos[74]" (29 casos), "exposições", "limãos" (cinco casos), "pavãos", "piãos", "piães", "religiães" e "uniãos" (quatro casos). Analisando essas palavras, observamos que todas elas são de freqüência baixa ou média, então pode existir um efeito da freqüência de ocorrência também nesse caso de variação nos plurais etimológicos em *–ões*. Organizamos a tabela abaixo, analisando o efeito da freqüência de ocorrência nas variações de plurais observadas nas palavras de etimologia *–ões*, ou seja, nos casos em que esses itens adotaram plural em *–ãos* ou *–ães*:

Tabela 24: Efeito da freqüência de ocorrência nas variações de plural etimológico *–ões*

Freqüência					
Baixa		**Média**		**Alta**	
N	**%**	**N**	**%**	**N**	**%**
41/43	95,3	02/43	4,6	0/43	0

Na tabela acima, observamos que, das 43 ocorrências de plurais etimológicos em *–ões* que adotaram plurais em *–ãos* e *–ães*, 41 (95,3%) foram com palavras de baixa freqüência e duas (4,6%) foram com palavras de freqüência média. Não houve nenhum caso de palavra etimológica em *–ões* com alta freqüência de ocorrência que adotasse plural em *–ãos* ou *–ães*. Observa-se, então, que as palavras de baixa e média

[74] Conforme afirmamos anteriormente, a palavra "bênção" tem etimologia em *–ões* (latim *benedictio*, *onis*). No entanto, por ser paroxítona, a pluralização prevista pela gramática normativa é "bênçãos". Esse é um caso raro de palavra etimológica em *–ões* que adota plural em *–ãos*, mas possivelmente isso ocorre por causa da regra que prevê que paroxítonas terminadas em *–ão* são sempre pluralizadas em *–ãos*.

freqüência de ocorrência são mais suscetíveis a mudanças de plural, mesmo na classe de *–ões*, que é o tipo mais freqüente do PB para as palavras terminadas em *–ão* no singular. No entanto, apesar de alguns itens com plurais etimológicos em *–ões* terem sido pluralizados em *–ãos* e *–ães*, essa migração é incipiente, atingindo apenas 6,3% do total dos dados (Cf. Tabela 23). Parece que, no cotidiano dos falantes, o uso dos plurais em *–ões* (tipo mais freqüente no léxico do PB) fixou formas que sofrem pequena taxa de variação. O Modelo de Redes e o de Exemplares explicariam essa baixa variabilidade em função da freqüência de tipo. Sendo o padrão em *–ões* altamente freqüente para a classe de itens terminados em *–ão* no singular, o grupo de palavras pluralizado através de *–ões* adquiriu exemplares mais fortes no léxico mental, o que dificulta a existência de variações nessa classe etimológica em *–ões*. Explica-se, portanto, por que migrações nesse grupo de plurais em *–ões* não ocorrem com a proeminência com que acontecem nos grupos de *–ãos* e *–ães* (tipos plurais menos freqüentes para palavras em *–ão* no singular).

Assim, através da análise etimológica das palavras em *–ão* (Cf. Tabela 23 acima), observamos que é pertinente a hipótese de que os plurais estão sendo generalizados na direção de *–ãos* ou *–ães* para *–ões*, e o contrário (de *–ões* para *–ãos* ou *–ães*) ocorre em menor amplitude. Em nossa pesquisa, estamos analisando justamente essas migrações de plural, por isso a exclusão de palavras cuja etimologia já é em *–ões* é necessária para que o SPSS possa selecionar de forma adequada as variáveis relevantes para a adoção de plurais analógicos. Se investigamos um grupo de itens em *–ãos* e *–ães* que migram para *–ões*, não faz sentido incluir, nessa mesma classe, palavras cujo plural etimológico já é em *–ões* e, conseqüentemente, não migrariam para aquilo que já são. Uma vez que definimos um objeto de estudo e uma variante a partir da qual os pesos relativos seriam calculados (no caso, *–ões*), não podemos incluir, em uma mesma análise no SPSS, palavras dos três grupos de plurais, já que *–ões* não pode se transformar naquilo que já é etimologicamente. Na verdade, a utilização de plurais etimológicos em *–ões* nos experimentos ocorreu apenas para verificar se haveria migração de *–ões* para *–ãos* ou *–ães*. Como isso aconteceu em poucos dados (apenas 43 em 670 ocorrências de plurais em *–ões*, ou seja, 6,3% do total) e já abordamos esses casos nos parágrafos anteriores, não é pertinente continuar com essas ocorrências em nossa análise. É possível, inclusive, que esses dados em *–ões* possam estar causando um mascaramento dos resultados relevantes para nossa pesquisa, elevando os percentuais de utilização do morfema *–ões* e ocultando possíveis resultados com relação a demais fatores que possam interferir na migração de plurais. Na verdade, até mesmo os

pesos relativos categóricos para freqüência de ocorrência (Cf. Tabela 21) podem ter sido causados pela interferência dos plurais em *–ões*, já que esses itens não podem ser contabilizados em um caso de migração que tem como alvo justamente o plural em *–ões*. A análise estatística está inadequada exatamente porque a inclusão de itens etimológicos em *–ões* impede o SPSS de selecionar as variáveis independentes relevantes para a migração de *–ãos* e *–ães* para *–ões* (já que este último é o fator a partir do qual os pesos relativos estão sendo calculados). A exclusão de plurais etimológicos em *–ões* nos dará, então, resultados mais transparentes quanto a migrações de plurais que ocorrem em direção à classe de *–ões*.

Deste ponto da pesquisa em diante, vamos excluir, portanto, os plurais etimológicos em *–ões* e analisar somente os dados de etimologia em *–ãos* e *–ães*, a fim de verificar a interferência de outros fatores nas generalizações de plural em direção a *–ões*. Ficamos, então, 12 palavras ("anão", "alemão", "cão", "cristão", "escrivão", "grão", "guardião", "irmão", "mão", "órgão", "pão", "vulcão"), que compõem 609 dados. Organizamos, primeiramente, outra tabela sobre etimologia, considerando apenas os plurais etimológicos em *–ãos* e *–ães* e analisando suas migrações em direção a *–ões*. Vejamos os resultados abaixo:

Tabela 25: Resultados gerais para palavras etimológicas em *–ãos* e *–ães*

Formas de Plural	**N**	**%**
Plural em *–ões*	178/609	29,2
Plural em *–ãos*	219/609	35,9
Plural em *–ães*	212/609	34,8

A tabela acima, portanto, não contém palavras de plural etimológico em *–ões*. Mesmo excluindo-se esse grupo, observamos que *–ões* ainda é adotado em 29,2% dos casos de palavras terminadas em *–ãos* e *–ães* etimológicos. As porcentagens de utilização dos outros dois morfemas (*–ãos* e *–ães*) são bem próximas, sendo que *–ãos* foi adotado em 35,9% dos casos e *–ães*, em 34,8%. Como na composição dessa tabela não havia dados de *–ões* etimológicos, não deveria haver tantos casos de adoção do plural em *–ões* (29,2%). Os resultados corroboram, mais uma vez, a hipótese de que as generalizações de plural ocorrem com as formas *–ãos* e *–ães* (tipos infreqüentes) migrando para o grupo de *–ões* (tipo mais freqüente). Os dados apresentados na tabela acima refletem melhor as variações nas classes de plurais, uma vez que consideramos apenas aqueles casos que efetivamente são

sujeitos a migrações de plurais em direção a *–ões*, excluindo, portanto, as ocorrências de *–ões* etimológicos, visto que esses não poderiam migrar para a classe a que já pertencem. Observa-se, então, que, conforme o Modelo de Redes postula, a freqüência de tipo é a principal responsável por generalizações analógicas, uma vez que um padrão lingüístico recorrente atrai mais membros para sua classe. Como *–ões* apresenta a freqüência de tipo mais alta para as palavras terminadas em *–ão* no singular, os plurais etimológicos em *–ãos* e *–ães* estão migrando para essa classe.

Prosseguindo com nossa análise, submetemos novamente os dados (sem os plurais etimológicos em *–ões*) ao programa SPSS. Nesse caso, as variáveis selecionadas foram **palavra** e **faixa etária**. Os resultados para a variável **palavra**, considerando somente os itens etimológicos em *–ãos* e *–ães*, são:

Tabela 26: Efeito da palavra nos plurais em *–ão*

Palavras	**Freqüência**	**N**	**%**	**P.R.**
1. Escrivão	baixa – 14	17/35	48,5	0.99
2. Guardião	baixa – 26	15/35	42,8	0.99
3. Vulcão	baixa – 62	101/102	99	0.99
4. Pão	**média – 118**	**0/106**	**0**	**0**
5. Anão	média – 142	24/36	66,6	0.99
6. Cristão	média – 235	12/35	34,2	0.99
7. Cão	**média – 350**	**0/35**	**0**	**0**
8. Grão	**média – 471**	**0/14**	**0**	**0**
9. Alemão	alta – 1.025	5/36	13,8	0.99
10. Irmão	**alta – 1.125**	**0/35**	**0**	**0**
11. Órgão	alta – 1.578	4/36	11,1	0.99
12. Mão	**alta – 2.949**	**0/104**	**0**	**0**

Na tabela acima, observamos que houve cinco palavras que não sofreram nenhum caso de migração de plurais. Esses itens são: "pão", "cão", "grão", "irmão" e "mão" (em negrito). Todas essas palavras são de freqüência de ocorrência média ou alta. Nenhuma delas pertence à freqüência baixa de ocorrência. Observamos, portanto, que o fato de um item léxico ser freqüente é um dos grandes responsáveis pela manutenção de

seu plural etimológico. Por outro lado, percebemos, também, que, nesse grupo de palavras (que não sofreram variação de plural), quatro são monossílabas[75].

A tabela acima apresenta, para todas as palavras, pesos relativos de zero ou 0.99 quanto à adoção do plural em *–ões*. Esse não é um resultado probabilístico que viabilize conclusões precisas sobre os itens léxicos que favorecem a migração para o plural em *–ões*. Provavelmente, esse resultado categórico ocorreu por interferência dos dados que não sofreram nenhum caso de migração de plurais para *–ões* (comentados no parágrafo anterior). Sendo assim, para obter resultados probabilísticos mais graduais, faz-se necessário remover de nossa análise todos os dados que categoricamente não tiveram nenhuma aplicação de plural em *–ões*. Portanto, as palavras "pão", "cão", "grão", "irmão" e "mão" foram excluídas de nosso banco de dados. Ficamos, então, com sete palavras ("alemão", "anão", "cristão", "escrivão", "guardião", "órgão" e "vulcão") e um total de 315 dados. Submetemos novamente essas ocorrências ao programa SPSS, a fim de verificar com mais precisão os pesos relativos de migração de plurais para cada palavra em particular. Os bancos de dados considerados nos dois casos (com todas as palavras e excluindo aquelas que não sofreram nenhum caso de migração de plural) são diferentes, por isso não se pode dizer que os resultados são totalmente comparáveis. No entanto, como é necessário obter medidas mais acuradas quanto às probabilidades de migração de plurais, precisamos excluir os dados das palavras já mencionadas anteriormente, apesar de estarmos cientes de que trabalhamos, nos dois casos, com universos distintos de dados. Os resultados para a variável palavra, excluídos os casos mencionados anteriormente, são os seguintes:

Tabela 27: Reanálise do efeito da palavra nos plurais em *–ão*

Palavras	**Freqüência**	**N**	**%**	**P.R.**
1. Escrivão	baixa – 14	17/35	48,5	0.47
2. Guardião	baixa – 26	15/35	42,8	0.40
3. Vulcão	**baixa – 62**	**101/102**	**99**	**0.99**
4. Anão	**média – 142**	**24/36**	**66,6**	**0.69**
5. Cristão	média – 235	12/35	34,2	0.30
6. Alemão	alta – 1.025	5/36	13,8	0.11
7. Órgão	alta – 1.578	4/36	11,1	0.08

[75] Ainda no decorrer desta análise sobre plurais em *–ão*, vamos comentar mais detalhadamente o fato de que os monossílabos costumam bloquear a migração em direção ao plural *–ões*.

Com a nova distribuição dos dados, os efeitos da variável palavra podem ser analisados com mais precisão. Observamos que os líderes nas migrações de plural foram os itens "vulcão" e "anão" (em negrito na tabela acima), sendo o primeiro com baixa freqüência de ocorrência e o segundo, com freqüência média. Percebemos, então, que a freqüência de ocorrência interfere na suscetibilidade do item léxico à mudança lingüística, fazendo com que itens mais recorrentes sejam menos propensos a mudanças na classe de plural etimológico e palavras menos freqüentes sejam as primeiras a migrar em direção a um paradigma mais freqüente (em *–ões*). Maiores discussões sobre o efeito da freqüência de ocorrência nas variações dos plurais em *–ão* serão discutidos adiante, ainda nesta análise. Não mencionaremos os casos menos propensos às migrações de plurais, porque, na Tabela 26, já comentamos os itens que não adotaram nenhum plural analógico em *–ões*. Aquelas palavras foram, portanto, as retardatárias no processo de migração de plurais em *–ão*.

Prosseguindo com nossa análise, devemos mencionar que, em nosso corpus, a variável palavra era bastante estratificada em fatores, já que contabilizamos 12 itens em *–ãos* e *–ães* etimológicos. Considerando-se que havia um total geral de 609 ocorrências (sem os etimológicos em *–ões*), seria esperado que essa variável fosse selecionada pelo SPSS como relevante, devido ao grande fatiamento do fator em relação ao número de dados. Por outro lado, quando consideramos variáveis como número de sílabas, tonicidade, estrutura morfológica, etc., existe uma sobreposição de fatores com relação à palavra, já que um mesmo item léxico apresenta, concomitantemente, um determinado número de sílabas, tonicidade e estrutura morfológica. Sendo assim, como já fizemos os comentários necessários acerca das palavras mais e menos propensas às migrações de plural, vamos, momentamente, excluir a variável palavra da análise, com o intuito de verificar se, sem ela, o SPSS selecionaria outros fatores igualmente significativos para as migrações em direção a *–ões*.

Retirando-se a variável palavra, o SPSS selecionou como relevantes os fatores **quantidade de plurais**, **número de sílabas**, **freqüência de ocorrência** e **faixa etária**[76]. Conclui-se, portanto, que, devido à alta variabilidade de códigos na variável palavra e por causa da interação com parâmetros lingüísticos (conforme explicamos anteriormente), a inserção do fator palavra estava impedindo que outras variáveis significativas para a análise

[76] O SPSS excluiu, portanto, as variáveis lingüísticas plural etimológico, tonicidade, estrutura morfológica e segmento precedente, e as variáveis extralingüísticas escolaridade e gênero. Tais fatores não serão retomados nesta análise porque não foram considerados significativos pelo programa SPSS. A única exceção que faremos é com relação à variável tonicidade, que, apesar de não ter sido selecionada pelo SPSS, será retomada ao longo desta análise.

fossem selecionadas pelo programa SPSS. Uma vez que já comentamos esse fator individualmente, podemos excluí-lo da análise e observar a interferência de outros fatores. Comentaremos cada uma das variáveis separadamente.

A primeira variável selecionada pelo SPSS foi **quantidade de plurais possíveis**. Estávamos investigando se o fato de algumas palavras em *–ão* apresentarem (de acordo com gramáticas do PB) um, dois ou até três plurais distintos interferiria na migração de plurais. Vejamos abaixo os resultados para esse fator.

Tabela 28: Efeito da quantidade de plurais possíveis nos plurais em *–ão*

Quantidade de Plurais	N	%	P.R.
Um Plural	39/437	8,9	0.33
Dois Plurais	15/35	42,8	0.18
Três Plurais	124/137	90,5	0.99

Os resultados da tabela acima mostram que, quando a palavra tem três possibilidades de plural (de acordo com as gramáticas normativas), há uma tendência à adoção do plural em *–ões* (peso relativo de 0.99). Quando há um ou dois plurais possíveis, *–ões* é desfavorecido (pesos relativos de 0.33 e 0.18, respectivamente). Esses resultados parecem comprovar que o morfema *–ões* é o mais disponível no léxico mental, então, quando a palavra tem, potencialmente, três plurais, *–ões* é o preferido. Pode-se argumentar que o fato de as palavras apresentarem uma, duas ou três possibilidades de plural é uma informação prevista pela gramática normativa do PB, não necessariamente real em termos de uso da língua. Consultando o Corpus NILC/São Carlos para verificar quais das palavras utilizadas em nossos experimentos realmente apresentavam uma, duas ou três possibilidades de plurais no uso corrente da língua, observamos que essa crítica é pertinente. Organizamos a tabela abaixo, que lista as 12 palavras em *–ãos* e *–ães* etimológicos, e apresentamos suas quantidades de plurais possíveis (de acordo com gramáticas normativas do PB) e a freqüência plural no Corpus NILC/São Carlos para cada uma dessas possibilidades de pluralização[77]:

[77] As palavras estão organizadas em ordem crescente a partir da freqüência de ocorrência total das formas de plural.

Tabela 29: Correlação entre freqüência de ocorrência e quantidade de plurais possíveis para plurais em *–ão*

Palavras	**Quantidade de plurais possíveis**	**Freqüência plural no Corpus NILC/São Carlos**		
		–ãos	***–ães***	***–ões***
1. Escrivão	1	2	11	1
2. Guardião	2	xxx	13	13
3. Vulcão	3	0	0	62
4. Pão	1	xxx	118	xxx
5. Anão	3	0	0	142
6. Cristão	1	235	xxx	xxx
7. Cão	1	xxx	350	xxx
8. Grão	1	471	xxx	xxx
9. Alemão	1	xxx[78]	1.025	xxx
10. Irmão	1	1.125	xxx	xxx
11. Órgão	1	1.578	xxx	xxx
12. Mão	1	2.949	xxx	xxx

Com base na tabela acima, observamos que, quando a gramática normativa registra que uma palavra tem dois ou três plurais, essa informação não se ratifica, necessariamente, na análise de dados reais de língua. No Corpus NILC/São Carlos, em geral, o que se verifica é que somente uma dessas duas ou três formas de plural é efetivamente adotada. Os únicos casos de utilização de plurais diferentes para o mesmo singular são "escrivão" e "guardião". No caso de "escrivão", as gramáticas consultadas indicavam apenas o plural "escrivães", mas as formas "escrivãos" e "escrivões" foram, também, encontradas no Corpus NILC/São Carlos. Isso sugere que essa palavra já apresenta alguma variação na língua, mas as gramáticas ainda não registraram esse fato. Quanto ao item "guardião", algumas gramáticas já admitem dois plurais possíveis ("guardiões" e "guardiães"). Ambos os casos ocorreram no Corpus NILC/São Carlos.

Portanto, os dados da tabela acima demonstram que o uso concreto da língua fixou palavras que só são adotadas com uma forma de plural, independentemente de a gramática normativa prescrever três possibilidades diferentes. No entanto, se as gramáticas listam esses plurais abundantes, é possível que seu uso seja atestado em algum nível, especialmente em situações formais escritas de uso da linguagem. Devemos lembrar que o Corpus NILC/São Carlos, conforme vimos no Capítulo 5, "Metodologia", é composto por

[78] Nesta tabela, o símbolo "xxx" indica que a palavra mencionada na primeira coluna não apresenta essa forma específica de pluralização.

dados de escrita. Em função disso, é possível que sofra a influência de revisores de texto, o que certamente minimiza a ocorrência de formas não-padrão, com os plurais em *–ões* analógicos. Ainda assim, casos de plurais divergentes do etimológico, como "escrivões" e "guardiões", foram encontradas para os itens terminados em *–ão*, o que demonstra, mais uma vez, que palavras em *–ãos* e *–ães* estão migrando para *–ões*.

A seguir, apresentamos os resultados relativos ao **número de sílabas**, outro fator considerado relevante pelo SPSS. Nessa variável, consideramos se a palavra era monossílaba, como "mão" e "pão", ou polissílaba, como "escrivão" e "órgão". Vejamos os resultados abaixo:

Tabela 30: Efeito do número de sílabas nos plurais em *–ão*

Número de Sílabas	N	%	P.R.
Monossílabos	0/260	0	0
Polissílabos	178/349	51	0.99

Com relação ao número de sílabas, observamos que não houve nenhum caso de item monossílabo a ser pluralizado com o morfema *–ões*. É importante lembrar que, no léxico do PB, não há itens monossílabos pluralizados em *–ões*. Todas as palavras monossílabas com singular em *–ão* formam seu plural em *–ãos* ou *–ães*, como "grãos" e "pães". Então, os falantes preservaram essa informação quando pluralizaram as palavras monossílabas. Segundo o Modelo de Redes, palavras inteiras são estocadas no léxico mental, portanto as formas de singular e de plural estão interligadas e informações morfológicas emergem dessas redes, levando os falantes a preservarem a pluralização etimológica de itens monossílabos. Nesse caso específico, pode-se observar, também, a presença de um "esquema" local (Cf. BYBEE; SLOBIN, 1982), que faz com que palavras de uma só sílaba tenham peculiaridades em sua representação lexical que as previnam de sofrer alterações em seus plurais. No léxico mental, as similaridades fonológicas entre palavras terminadas em *–ão* fazem com que elas componham uma rede. Informações morfológicas, como o fato de que todos os monossílabos são pluralizados em *–ãos* ou *–ães*, derivam dessas redes, daí o fato de nenhuma palavra de uma sílaba ter sido pluralizada em *–ões*. Apesar de *–ões* ser o tipo de plural mais freqüente para a classe de *–ão* no singular, o grupo de palavras monossílabas está, de certa forma, imune a essa variação, devido à sua forte representação no léxico mental.

O último fator lingüístico considerado relevante pelo SPSS foi **freqüência de ocorrência**. As possibilidades eram freqüência baixa (até 99 ocorrências), média (de 100 a 500) e alta (mais de 500). Vejamos os resultados a seguir:

Tabela 31: Efeito da freqüência de ocorrência nos plurais em *–ão*

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	133/172	77,3	0.84
Freqüência Média	36/227	15,8	0.45
Freqüência Alta	9/210	4,2	0.18

Para melhor visualizarmos os resultados sobre freqüência, vamos reorganizar os dados da tabela acima no gráfico abaixo:

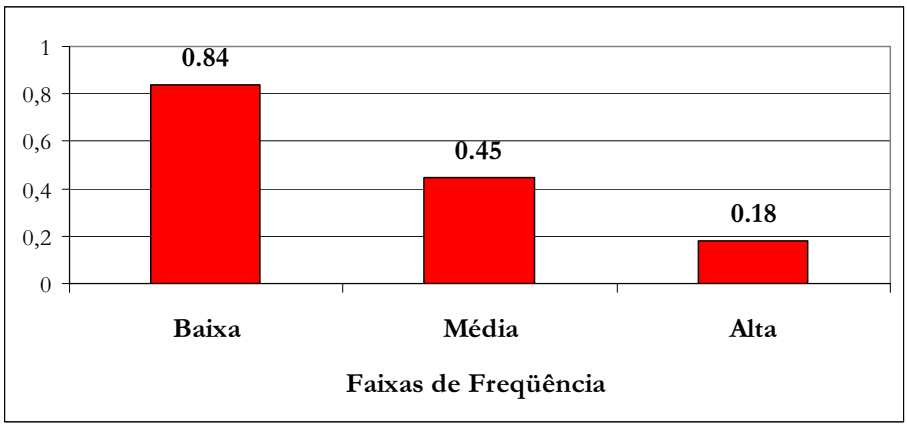


Gráfico 1: Efeito da freqüência de ocorrência nos plurais em *–ão*

Na tabela e no gráfico acima, observamos que houve uma gradação na adoção de plurais em *–ões*, das palavras mais freqüentes às menos freqüentes. O gráfico, particularmente, permite uma visualização mais clara da distribuição dos dados de acordo com a freqüência de ocorrência. As palavras de baixa freqüência favorecem o plural em *–ões* (0.84), ao passo que as de freqüência média desfavorecem (0.45) e as de freqüência alta desfavorecem mais ainda (0.18). Os resultados comprovam, portanto, que os falantes preservam mais os plurais etimológicos nas palavras freqüentes. Isso sugere que a forma como as palavras são usadas no cotidiano afeta a organização das informações lingüísticas no léxico mental. Na verdade, a representação mental é modelada e remodelada a partir do uso da língua. Quanto mais uma palavra é usada, mais ela se torna disponível no léxico

mental, facilitando, assim, seu acesso e prevenindo-a, de certa maneira, contra mudanças analógicas. Para o Modelo de Redes, a interação entre freqüência de tipo e de ocorrência é responsável pela organização das representações mentais. Assim, quando a palavra tem alta freqüência de ocorrência, torna-se autônoma no léxico mental e não depende de sua classe paradigmática para que sua forma seja recuperada. Se essa palavra pertencer a um tipo pouco freqüente no léxico, tende a manter esse padrão, porque sua alta freqüência de ocorrência faz com que ela se torne mais forte e resista a mudanças analógicas; por outro lado, se a palavra tem baixa freqüência de ocorrência e pertence a um padrão (tipo) também pouco freqüente, acaba por se submeter ao paradigma analógico mais recorrente, no caso, *–ões*. É isso o que observamos nos plurais em *–ão*: os tipos infreqüentes em *–ãos* e *–ães* foram mantidos apenas nas palavras freqüentes, que não dependem de sua classe para que sejam acessadas; palavras de baixa e média freqüência de ocorrência favoreceram o plural em *–ões*, já que esse é o mais produtivo para palavras terminadas em *–ão* no singular e não há, também, restrições fonológicas para que ele se aplique a qualquer item de forma singular em *–ão*. Esses resultados para freqüência de ocorrência são compatíveis com nossa hipótese de trabalho.

A Teoria de Palavras e Regras postula efeitos de freqüência de ocorrência apenas para itens irregulares. Se os regulares não estão estocados no léxico mental e são derivados a partir de regras, não devem ser suscetíveis a efeitos de freqüência de ocorrência. Itens irregulares, por sua vez, podem ser influenciados por freqüência, já que sua recuperação no léxico mental depende de memorização, não de regras. No caso sob análise, observa-se que as palavras pluralizadas em *–ãos* seriam regulares (flexionadas a partir do morfema *–s*, o canônico para plurais do PB) e as pluralizadas em *–ães* seriam irregulares. Esperar-se-ia, portanto, em consonância com os postulados da Teoria de Palavras e Regras, que apenas palavras pluralizadas em *–ães* fossem suscetíveis a efeitos de freqüência de ocorrência. No entanto, nossos dados mostram que itens em *–ãos* e *–ães* foram igualmente afetados por freqüência de ocorrência, já que palavras desses dois plurais etimológicos migraram para a classe de *–ões*, como, por exemplo, "cristões" (em vez de "cristãos") e "guardiões" (em vez de "guardiães"). A Teoria de Palavras e Regras, portanto, não seria capaz de justificar por que tanto palavras regulares quanto irregulares apresentaram efeitos de freqüência de ocorrência.

A seguir, vamos comentar, brevemente, o efeito da **tonicidade** para os plurais em *–ão*. Conforme mencionamos anteriormente, esse fator não foi selecionado na análise

do SPSS. No entanto, em nossos dados havia apenas duas palavras paroxítonas em *–ão*: "bênção" e "órgão". Como vimos, "bênção" é de etimologia *–ões* e, por isso, foi excluída da análise. Ficamos, portanto, somente com um item paroxítono ("órgão") e, por isso, não havia quantidade suficiente de dados para que o fator tonicidade fosse selecionado pelo SPSS. No entanto, já que a gramática normativa do PB prevê que todas as palavras paroxítonas são pluralizadas em *–ão*, consideramos que seria importante verificar se os falantes preservariam essa informação quando provessem o plural desses itens. Assim, organizamos a tabela abaixo, contendo apenas os resultados sobre tonicidade, considerando as palavras "bênção" e "órgão". Vejamos os resultados:

Tabela 32: Efeito da tonicidade nos plurais em *–ão*

Tonicidade	**N**	**%**
Oxítonas	174/573	30
Paroxítonas	10/71	14

Com relação à tonicidade, observamos que as palavras oxítonas adotaram plural em *–ões* em 30% dos casos, ao passo que as paroxítonas adotaram esse mesmo plural em uma porcentagem menor dos dados (14%). Em nossos experimentos, os informantes tenderam a preservar a informação de que os plurais de palavras paroxítonas (segundo a gramática normativa) são em *–ãos*. Esse resultado sugere que os falantes armazenam palavras inteiras no léxico mental e que generalizações morfológicas emergem a partir das redes de conexões semânticas e fonológicas estabelecidas entre as palavras no léxico mental. Portanto, a informação sobre tonicidade para itens terminados em *–ão* decorre de conexões entre palavras inteiras estocadas no léxico mental. Observa-se, então, que, dentro da classe de plurais em *–ão*, há uma generalização local que, até certo ponto, previne palavras paroxítonas de sofrerem analogia baseada na freqüência de tipo em *–ões*. No entanto, mesmo nesse grupo de paroxítonas, percebe-se que já houve dez ocorrências de adoção dos plurais "bênções" e "órgões", o que demonstra a força do paradigma em *–ões*, decorrente da sua alta freqüência de tipo.

Quanto aos fatores extralingüísticos, o único que o SPSS considerou relevante foi **faixa etária**, estratificada da seguinte maneira: jovens (de 15 a 20 anos), medianos (de 35 a 40) e adultos (de 55 a 60 anos). Vejamos abaixo os resultados para essa variável.

Tabela 33: Efeito da faixa etária nos plurais em *–ão*

Faixas Etárias	N	%	P.R.
Jovens	75/208	36	0.74
Medianos	61/206	29,6	0.51
Adultos	42/195	21,5	0.24

Os números da tabela acima demonstram que existe uma distribuição dos dados compatível com mudança em progresso (em termos labovianos[79]), já que os jovens tendem a adotar as formas inovadoras (com plurais em *–ões*) mais que os adultos (peso relativo de 0.74 para os jovens, contra 0.24 para os adultos). Observamos que a faixa etária mais alta foi a que mais preservou os plurais etimológicos em *–ãos* ou *–ães*. Nossa interpretação para esses resultados é de que o léxico mental dos falantes mais idosos ainda preserva os plurais etimológicos, mas a força do tipo em *–ões* é tanta, que os falantes mais jovens estão aplicando esse paradigma quando querem pluralizar as palavras terminadas em *–ão* no singular[80]. Um contra-argumento a essa hipótese é que, se o léxico mental das pessoas é atualizado constantemente, poderíamos esperar que tanto falantes jovens quanto adultos armazenassem e utilizassem formas generalizadas em *–ões*. No entanto, conforme afirmamos anteriormente, o léxico mental armazena itens inteiros e até mesmo informações pragmáticas associadas ao uso das palavras. Assim, pode ser que os falantes mais velhos tenham armazenado formas generalizadas no léxico mental, mas evitem adotá-las, por saberem que essas não são as formas consideradas "corretas" pela norma padrão do português.

Ainda com relação à faixa etária, é importante lembrar, também, que a ausência de marca formal de pluralização ("os guardiãoØ") é bastante recorrente no PB. Poder-se-ia hipotetizar, então, que os jovens favorecem os plurais em *–ões*, e lideram, também, o processo de cancelamento das marcas formais de plural. No entanto, em nossos dados totais sobre os plurais em *–ão*, houve 62 ocorrências de apagamento das marcas formais de plural, o que corresponde a 6,4% do total de palavras coletadas. Desses casos, 20 (32,2%) ocorreram entre os jovens, 15 (24,1%) entre os medianos e 27 (43,5%) entre os adultos. Não encontramos, portanto, nenhum padrão que demonstre que a perda da marcação de

[79] A discussão sobre a Sociolingüística colocar-nos-ia além da proposta de análise teórica de nossa tese. Para maiores informações sobre esse assunto, consulte-se Labov (1972, 1994, 2001).
[80] Exceção deve ser feita quanto aos itens monossílabos, porque, como vimos na Tabela 30, esses itens não sofrem variação.

plural é relacionada à faixa etária. No entanto, o número de dados é pequeno demais (62 ocorrências de apagamento de plural) para que qualquer afirmação categórica seja feita. A princípio, observamos que os jovens lideram o processo de migração de plurais, mas não estão à frente na perda das marcas formais de pluralização dos casos em *–ão*.

Finalizamos aqui nossas considerações sobre os plurais das palavras terminadas em *–ão* no singular. Na penúltima subseção deste capítulo retomaremos algumas informações sobre esse grupo, a fim de proceder à discussão geral dos dados. Na subseção seguinte, analisaremos os plurais das palavras terminadas em *–l* no singular.

6.2 Resultados para os plurais terminados em *–l*

Nesta subseção, dedicamo-nos à análise dos resultados para os grupos de plurais em *–l*, que contabilizaram 1.167 dados coletados em nossos experimentos. Para essa classe de plural, a variável dependente (resposta adotada pelos falantes) era constituída da seguinte forma: 1) Plural em *–l*; 2) Plural em ditongo em *–u*; 3) Ausência de plural: 4) Outro plural. Os resultados gerais para o grupo de plural em *–l* são os seguintes:

Tabela 34: Resultados gerais para os plurais em *–l*

Formas de Plural	N	%
Plural em *–l*	1.017/1.167	87,1
Plural em ditongo em *–u*	95/1.167	8,1
Ausência de plural	33/1.167	2,8
Outro plural	22/1.167	1,8

Dos 1.167 dados, observamos que 1.017 (87,1%) foram pluralizados através do morfema *–l*, o mais esperado, já que é assim que todas as palavras desse grupo terminam, mas houve 95 dados (8,1%) pluralizados através da regra de ditongo em *–u*[81]. Houve, também, casos de utilização do singular pelo plural, de forma similar à que ocorreu com os dados de *–ão*. Exemplos dessas ocorrências são: "Um conjunto de <u>móvel</u>" ou "Muitas

[81] Nos resultados das tabelas subseqüentes, vamos mencionar individualmente as palavras que sofreram migrações para a classe de ditongo em *–u*.

pessoas não gostam de ler jornal", casos em que o objeto na figura aparecia em número plural, mas os informantes utilizaram formas de singular. Ocorrências de outro plural foram: "cachecóiØ", "goles" (para "gol"), "meles" (para "mel"), "míssis" (quatro casos), "móvis" (11 casos), "sales" (para "sal") e "útis" (três casos). Dos 22 itens a adotarem outro plural, 18 (81,8%) eram paroxítonos. Parece que essas palavras estão instáveis na classe de plurais em *–l*, já que a maioria dos itens pluralizados em *–l* é oxítona. O ASPA lista 679 palavras do grupo em *–l* que são pluralizadas através da adição de *–is*. Dessa classe, 34 (5%) são paroxítonas e 645 (94,9%) são oxítonas. Já que há mais oxítonos que paroxítonos pluralizados em *–l*, pode estar havendo um efeito dessa freqüência de tipo na instabilidade dos itens paroxítonos. No entanto, os plurais divergentes adotados pelos falantes não transformam as palavras em oxítonas, já que, em "míssis", "móvis" e "útis", o acento continua na penúltima sílaba. Convém lembrar, também, que Nunes (1945, p. 229, conforme mencionado na subseção 2.3.1 do Capítulo 2, "Grupos de Plurais") comenta que, nos grupos em *–l*, a distinção de plural por tonicidade não existia no português arcaico. Sendo assim, palavras oxítonas e paroxítonas faziam plural da mesma maneira, através do morfema *–es* ou *–is*, em vez de *–eis*: "estavees", "notavees", "movis", "razoavis", em vez de "estáveis", "notáveis", "móveis", "razoáveis". Observamos, portanto, que casos semelhantes a "míssis", "móvis" e "útis" já eram registrados no português arcaico.

Feitos os comentários necessários sobre os casos de ausência de plural e adoção de outro plural para o grupo em *–l*, excluímos esses dados das rodadas seguintes do SPSS, a fim de fazer a regressão binária e alcançar resultados mais transparentes para a aplicação do plural em *–l* ou ditongo em *–u*. Como a ocorrência desses casos já foi comentada individualmente no parágrafo anterior, julgamos que maiores discussões sobre eles seriam redundantes. Ficamos, então, com um total de 1.112 ocorrências em que o plural em *–l* ou em ditongo em *–u* foi efetivamente adotado. Submetemos esses dados ao programa SPSS, a fim de verificar as variáveis relevantes para o processo de migração de plurais em *–l* para ditongo em *–u*.

O SPSS selecionou as variáveis independentes **palavra** e **escolaridade**. Os resultados para a variável **palavra** são apresentados abaixo[82]:

---

[82] Neste grupo de palavras terminadas em *–l*, são apresentados resultados relativos à aplicação do plural em ditongo em *–u*, já que, no SPSS, era necessário selecionar uma variante a partir da qual os pesos relativos seriam calculados. Como, nesse grupo, investigávamos se haveria migração de plurais em *–l* para ditongo em *–u*, definimos este último fator como indicativo do peso relativo. Portanto, quando a tabela apresenta, por exemplo, peso relativo de 0.99, esse valor refere-se à aplicação do plural em ditongo em *–u*.

Tabela 35: Efeito da palavra nos plurais em *–l*

Palavras	Freqüência	N	%	P.R.
1. Mel	baixa – 0	25/35	71,4	0.99
2. Cachecol	**baixa – 02**	**0/24**	**0**	**0**
3. Funil	baixa – 02	2/33	6	0.99
4. Anzol	**baixa – 08**	**0/36**	**0**	**0**
5. Avental	**baixa – 17**	**0/69**	**0**	**0**
6. Sol	baixa – 27	15/68	22	0.99
7. Gentil	**baixa – 31**	**0/36**	**0**	**0**
8. Pincel	**baixa – 37**	**0/34**	**0**	**0**
9. Pastel	**baixa – 63**	**0/36**	**0**	**0**
10. Agradável	baixa – 74	1/36	2,7	0.99
11. Farol	média – 100	1/49	2	0.99
12. Anel	**média – 110**	**0/101**	**0**	**0**
13. Lençol	**média – 118**	**0/36**	**0**	**0**
14. Sal	média – 129	7/35	20	0.99
15. Acessível	média – 165	1/36	2,7	0.99
16. Míssil	média – 237	3/23	13	0.99
17. Azul	média – 280	1/34	2,9	0.99
18. Infantil	**média – 328**	**0/36**	**0**	**0**
19. Espanhol	**alta – 501**	**0/35**	**0**	**0**
20. Difícil	alta – 654	2/35	5,7	0.99
21. Móvel	**alta – 728**	**0/53**	**0**	**0**
22. Útil	alta – 783	2/32	6,2	0.99
23. Sinal	**alta – 1.422**	**0/38**	**0**	**0**
24. Hospital	**alta – 1.513**	**0/36**	**0**	**0**
25. Responsável	**alta – 1.894**	**0/35**	**0**	**0**
26. Jornal	**alta – 2.490**	**0/56**	**0**	**0**
27. Gol	alta – 4.565	35/35	100	1

Na tabela acima, observamos que houve 15 palavras que não sofreram nenhuma migração de plural (em negrito). Esses itens são "cachecol", "anzol", "avental", "gentil", "pincel", "pastel", "anel", "lençol", "infantil", "espanhol", "móvel", "sinal", "hospital", "responsável" e "jornal". Nesse grupo, encontram-se palavras pertencentes às três faixas de freqüência de ocorrência. Não houve, portanto, predominância de palavras de uma freqüência específica (do grupo em *–l*) favorecendo os plurais em ditongo em *–u*. Devemos mencionar, também, o caso da palavra "gol", que foi pluralizada como ditongo

em *–u* em todas as suas ocorrências, obtendo, portanto, resultado categórico quanto à aplicação da variável dependente[83]. À parte esses itens (palavra "gol" e os que não sofreram nenhuma migração de plural), observamos que todos os demais apresentam peso relativo de 0.99. Provavelmente, como a maioria das palavras desse grupo (terminado em *–l*) não adotou plural analógico, esse resultado categórico está impedindo que medidas mais precisas de peso relativo para os demais itens apareçam. Assim, a fim de obter resultados mais adequados, removemos todos os dados referentes a palavras que não sofreram nenhuma migração de plural e excluímos, também, a palavra "gol", porque ela foi pluralizada como "gols" (ditongo em *–u*) em todas as suas ocorrências. Ficamos, portanto, com apenas 11 palavras (416 dados) em que houve, efetivamente, migração de plural do grupo de *–l* em direção à classe de ditongo em *–u*. Submetemos esses dados novamente ao programa SPSS e obtivemos números mais graduais de pesos relativos, conforme podemos visualizar na tabela abaixo:

Tabela 36: Reanálise do efeito da palavra nos plurais em *–l*

Palavras	Freqüência	N	%	P.R.
1. Mel	**baixa – 0**	**25/35**	**71,4**	**0.97**
2. Funil	baixa – 02	2/33	6	0.39
3. Sol	**baixa – 27**	**15/68**	**22**	**0.76**
4. Agradável	baixa – 74	1/36	2,7	0.21
5. Farol	média – 100	1/49	2	0.17
6. Sal	**média – 129**	**7/35**	**20**	**0.72**
7. Acessível	média – 165	1/36	2,7	0.21
8. Míssil	média – 237	3/23	13	0.67
9. Azul	média – 280	1/34	2,9	0.21
10. Difícil	alta – 654	2/35	5,7	0.36
11. Útil	alta – 783	2/32	6,2	0.42

Após a exclusão dos dados de efeito categórico para a variável palavra, observamos com mais precisão que os itens "mel", "sal" e "sol" (em negrito acima) favoreceram a migração de plurais da classe em *–l* para ditongo em *–u*. Como característica comum a essas palavras, apontamos o fato de todas serem monossílabas. Conforme veremos posteriormente, ainda no decorrer desta análise, itens de uma só sílaba costumam

[83] Na discussão dos resultados para vogal precedente, comentaremos maiores detalhes sobre a palavra "gol".

favorecer os plurais em ditongo em *–u*. Por outro lado, "mel" e "sol" são palavras de baixa freqüência de ocorrência e "sal", de média. Observamos, portanto, que as palavras pouco freqüentes são, provavelmente, menos acessíveis no léxico mental e, conseqüentemente, mais suscetíveis a mudanças na classe de pluralização. A freqüência de ocorrência para itens de plural em *–l* será comentada posteriormente, no desenrolar desta análise.

Conforme já havíamos mencionado no caso dos plurais em *–ão*, a estratificação dos dados por palavra era ampla (27 itens terminados em *–l*), portanto era esperado que essa variável fosse significativa. Além disso, havia, também, sobreposição de fatores lingüísticos com relação às palavras (de forma análoga à que mencionamos na subseção anterior), por isso excluímos a variável palavra e fizemos novas rodadas no SPSS. Sem esse fator, o SPSS selecionou as variáveis **estrutura morfológica**, **vogal precedente**, **número de sílabas** e **escolaridade**[84]. Comentaremos cada uma delas em separado.

Analisaremos, primeiramente, **estrutura morfológica**. As opções eram: palavras com sufixo, como "agradável" ou "gentil" e palavras sem sufixo, como "azul" ou "útil". Os resultados são apresentados na tabela abaixo:

Tabela 37: Efeito da estrutura morfológica nos plurais em *–l*

Estrutura Morfológica	N	%	P.R.
Sufixo	2/336	0,59	0
Não-Sufixo	93/776	11,9	0.99

A tabela acima demonstra que os itens que não têm sufixo favorecem a migração para a classe de ditongo em *–u*. No entanto, os pesos relativos apresentados (zero para itens com sufixo e 0.99 para palavras sem sufixo) são categóricos e parecem não refletir de forma apropriada a realidade dos números e porcentagens de variações de plural nesse grupo de palavras terminadas em *–l*. Especialmente para os itens sem sufixo, o valor de peso relativo em 0.99 parece alto demais para uma distribuição percentual de 93 em 776 (11,9%). Portanto, é provável que a presença de algumas variáveis esteja impedindo que o programa alcance pesos relativos mais precisos quanto ao efeito da estrutura morfológica[85].

[84] O SPSS excluiu, portanto, as variáveis lingüísticas tonicidade e freqüência de ocorrência, bem como as variáveis extralingüísticas faixa etária e gênero. Não faremos mais comentários acerca desses fatores, com exceção de freqüência de ocorrência, que será retomada ao longo desta análise, ainda nesta subseção.

[85] O SPSS é um programa minucioso de análise estatística, no entanto a interpretação dos pesos relativos por ele obtidos deve ser cautelosa, observando se houve interação entre variáveis que possam causar resultados incompatíveis entre os pesos relativos e os cálculos percentuais. Nos casos em que observamos a ocorrência

Investigando por que isso ocorreu, excluímos de nossa análise as variáveis lingüísticas vogal precedente e número de sílabas, porque, em nossos dados, os itens com vogal precedente [o] e [u] e todas as palavras monossílabas não apresentavam sufixo. Havia, portanto, lacunas nos dados que estavam fazendo com que as variáveis vogal precedente e número de sílabas (também relevantes para a análise) interagissem com estrutura morfológica, impedindo, assim, o cálculo de pesos relativos mais adequados aos números percentuais. Excluindo-se esses fatores, os pesos relativos são menos categóricos e refletem, de forma mais apropriada, a distribuição numérica dos dados:

Tabela 38: Reanálise do efeito da estrutura morfológica nos plurais em *–l*

Estrutura Morfológica	**N**	**%**	**P.R.**
Sufixo	2/336	0,59	0.16
Não-Sufixo	93/776	11,9	0.83

Observamos, através desta tabela, que a estrutura morfológica é, realmente, um fator determinante para que haja migrações de plural *–l* em direção a ditongo em *–u*. Os resultados demonstram que palavras que não têm sufixo favorecem a variação de plural (0.83) e palavras com sufixo a desfavorecem (0.16). Provavelmente, esse resultado ocorreu porque, no grupo de ditongo em *–u*, existem apenas dois sufixos (*–eu* e *–aréu*), ao passo que, no grupo em *–l*, Coutinho (1981) e Cunha e Cintra (1985) listam oito sufixos diferentes (Cf. Capítulo 2, "Grupos de Plurais"). Sendo assim, a existência de sufixo parece ser uma forte referência para que o falante preserve a forma de plural padronizada em *–l* (segundo a gramática normativa do PB). Mesmo assim, identificamos dois dados em que itens terminados em *–l* com sufixo foram pluralizados através da regra de ditongo em *–u*: "acessívels" e "agradávels". Os itens que não apresentavam sufixo e foram pluralizados na regra de ditongo em *–u* foram: "azuls", "difícils" (dois casos), "farols", "funils", (dois casos), "gols" (35 casos), "mels" (25 casos), "míssils" (três casos), "sals" (sete casos), "sols" (15 casos)" e "útils" (dois casos). Nesse grupo, observamos que as palavras monossílabas são as líderes na migração para o plural em ditongo em *–u*. Discutiremos o fator número de sílabas nos parágrafos subseqüentes.

---

de tais interações, removemos algumas variáveis, a fim de observar o melhor ajuste para os valores de peso relativo. Todas essas exclusões são devidamente comentadas ao longo deste capítulo.

Ainda com relação à estrutura morfológica dos itens em *–l*, é interessante que palavras que tenham sufixo sejam mais resistentes à migração para a classe de plurais em ditongo em *–u*. Mais uma vez, observamos a interferência da freqüência de tipo, de acordo com os postulados da Teoria dos Exemplares e do Modelo de Redes. Já que a estrutura morfológica emerge das relações semânticas e fonológicas entre as palavras armazenadas no léxico mental dos falantes, podemos hipotetizar que a informação sobre sufixo é uma forte referência para que uma palavra preserve sua pluralização em *–l*. Há indícios, portanto, de que o léxico mental preserva informações probabilísticas sobre a freqüência de ocorrência das estruturas lingüísticas. Sendo sufixação uma das características marcantes do grupo de palavras terminadas em *–l*, os falantes acabam por preservar as formas padronizadas de plurais em *–l* quando as palavras têm sufixo.

O próximo fator considerado relevante na análise dos plurais em *–l* foi **vogal precedente**. As possibilidades eram: vogal precedente [a] ("jorn**a**l"), [ɛ] ("past**e**l"), [e] ("agradáv**e**l"), [i] ("fun**i**l"), [ɔ] ("anz**o**l"), [o] (o único item era "g**o**l") e [u] (o único item era "az**u**l"). Vejamos os resultados na tabela abaixo:

Tabela 39: Efeito da vogal precedente nos plurais em *–l*

Vogal Precedente	N	%	P.R.
[a]	7/234	2,9	0
[ɛ]	25/205	12,1	0.003
[e]	**2/160**	**1,2**	**0.99**
[i]	9/195	4,6	0.06
[ɔ]	16/249	6,4	0
[o]	**35/35**	**100**	**0.99**
[u]	1/34	2,9	0.02

Na tabela acima, observamos que as vogais precedentes [e] ou [o] (em negrito) favorecem as migrações de plurais em *–l* na direção de ditongo em *–u*. No entanto, esses resultados devem ser analisados com cautela, visto que, nesta tabela, os pesos relativos também parecem não corresponder à distribuição numérica. Percebemos, por exemplo, que, nos dois casos que favorecem a variação de plural (vogal precedente [e] ou [o]) os resultados de número e porcentagem são completamente diferentes: i) Para [e],

houve dois casos de migração em 160, o corresponde a 1,2% dos dados; para [o], houve aplicação categórica de 35 migrações em 35 palavras, ou seja, 100%. Os valores numéricos e percentuais são, portanto, bastante diferentes, no entanto os dois pesos relativos são iguais (0.99). Esse número parece não ser adequado à distribuição dos dados, por isso, a exemplo do que fizemos na tabela anterior, resolvemos investigar possíveis interferências entre as variáveis selecionadas pelo SPSS. Observamos que havia problemas, primeiramente, entre vogal precedente e número de sílabas, porque os itens terminados em vogal [e], [i] ou [u] não apresentavam itens monossílabos; outra influência constatada foi entre vogal precedente e estrutura morfológica, porque os itens terminados em [o] ou [u] ("gol" e "azul") eram sem sufixo. Assim, removemos as variáveis número de sílabas e estrutura morfológica, para que os pesos relativos para as vogais precedentes pudessem ser mais compatíveis com os percentuais apresentados. Excluídos esses fatores, os novos resultados para vogal precedente são:

Tabela 40: Reanálise do efeito da vogal precedente nos plurais em *–l*

Vogal Precedente	N	%	P.R.
[a]	7/234	2,9	0.03
[ɛ]	25/205	12,1	0.09
[e]	2/160	1,2	0
[i]	9/195	4,6	0.01
[ɔ]	16/249	6,4	0.04
[o]	35/35	100	1
[u]	1/34	2,9	0.05

Excluindo-se as variáveis lingüísticas estrutura morfológica e número de sílabas, os pesos relativos são mais compatíveis com os números e percentuais alcançados. Observamos, por exemplo, que, para a vogal [e], o peso relativo referente a 1,2% agora é zero e, para a vogal [o], o peso relativo para 100% agora é 1. Portanto, sem essas variáveis (estrutura morfológica e número de sílabas), podemos analisar melhor a interferência da vogal precedente na migração de plurais em *–l* para ditongo em *–u*.

Na nova tabela acima, todos os pesos relativos para as vogais precedentes são bastante próximos, à exceção do resultado para a vogal [o], que é categórico, com a

aplicação de 100% de plural em ditongo em *–u* para palavras terminadas em *–l*. Conforme afirmamos anteriormente, em nossos experimentos, o único item que apresentava vogal final [o] era a palavra "gol", portanto pluralizada como "gols" (ditongo em *–u* em vez de *–l*) em todas as suas 35 ocorrências. Então, houve um item léxico favorecendo a adoção de plural em ditongo em *–u*, não uma vogal específica que fosse mais propensa à migração de plurais.

Com relação à palavra "gol", algumas considerações devem ser feitas. Em geral, conforme afirmamos na subseção 3.3 do Capítulo 3, "Revisão de Literatura", é esperado que itens terminados em ditongo em *–u* adotem a regra de pluralização em *–l*, já que este tipo é mais freqüente que aquele. Exemplos disso são variações de plurais como "chapéis", "mausoléis" e "troféis", todas presentes em nossos dados. Por causa da baixa freqüência do tipo ditongo em *–u* no PB, não esperávamos que houvesse muitos itens migrando da classe em *–l* para a de ditongo em *–u*. O item "gol" parece contradizer isso. Essa palavra foi importada do inglês "goal" e sua grafia oficial no PB suprimiu a letra "a", a fim de que a palavra se adequasse aos padrões ortográficos da nossa língua. No entanto, apesar de a palavra terminar em *–l*, sua pluralização nas transmissões de jogos de futebol, eventos em que ela é bastante freqüente, geralmente é [ˈgows]. Esse fato demonstra justamente o contrário do que mencionamos anteriormente sobre generalizações de plural. Nesse caso, está sendo usado o plural de ditongo em *–u* (acréscimo de *–s*) para uma palavra que termina em *–l* (a regra esperada seria a supressão de *–l* e o acréscimo de *–is*, produzindo a forma [ˈgoys]). Explicações possíveis para a pluralização dessa palavra são: 1) No inglês, língua originária desse item, seu plural é formado com o acréscimo do morfema *–s*, então o plural de "goal" é "goals". É possível que o PB tenha adotado, simultaneamente, as formas singular e plural, mantendo, nos dois casos, grafias similares às originárias; 2) Pode ser que "gol" tenha sido importado para o PB quando o processo de vocalização de *–l* em fim de sílaba já estava avançado, portanto a palavra talvez já tenha entrado na língua como [ˈgow]. Consultamos o Dicionário Houaiss on-line, disponível em <http://houaiss.uol.com.br/busca.jhtm> e observamos que o ano de 1904 é o primeiro registro da palavra "gol" no português escrito. Não sabemos, no entanto, se a vocalização do *–l* em fim de sílaba já era um fenômeno recorrente no PB dessa época. Nenhuma menção sobre os primeiros registros de vocalização do *–l* foi encontrada nas gramáticas históricas pesquisadas. O item "gols" é, portanto, um caso raro em que uma palavra terminada em *–l* adota a regra de plural em *–u*.

Prosseguindo com nossa análise, vamos comentar a variável **número de sílabas**, cujas possibilidades consideradas eram monossílabos, como "mel" e "sal", e polissílabos, como "cachecol" e "farol". Os resultados são apresentados na tabela a seguir:

Tabela 41: Efeito do número de sílabas nos plurais em *–l*

Número de Sílabas	N	%	P.R.
Monossílabos	82/173	47,3	0.96
Polissílabos	13/939	1,3	0.03

A tabela acima mostra que monossílabos favorecem a pluralização de ditongo em *–u* (peso relativo de 0.96) e os polissílabos desfavorecem (0.03). Esse resultado é contrário àquele apresentado para os plurais em *–ão*, já que, naquele grupo, não houve nenhum monossílabo em *–ãos* ou *–ães* que migrasse para a classe de *–ões*. Pode ser que a baixa freqüência de ocorrência das palavras monossílabas tenha feito com que elas migrassem para a regra de ditongo em *–u*. Os itens monossílabos que adotaram plural em ditongo em *–u* foram: "gol" (35 casos), "mel" (25 casos), "sal" (sete casos) e "sol" (15 casos)". Observamos que, dessas três palavras, "mel" e "sol" têm freqüência baixa, "sal" tem freqüência média e "gol" tem freqüência alta, mas essa é uma palavra atípica, conforme justificamos anteriormente. Por outro lado, como vimos no Capítulo 3, "Revisão de Literatura", o tipo em *–l* é muito mais freqüente no PB que o tipo em *–u*, portanto não seria esperado que palavras pluralizadas etimologicamente em *–l* adotassem plural em *–u*.

Uma explicação possível para o fato de os itens monossílabos favorecerem a aplicação do plural em ditongo em *–u* é a glidificação que ocorre no português falado no estado do Rio de Janeiro. Abaurre (1973, p. 14) comenta o caso de uma regra de inserção de glide no PB (em casos de vogal tônica final seguida de *–s*), o que faz com que, por exemplo, o plural de "bom" seja [ˈbõỹs] e o de "fim" seja [ˈfĩỹs], ambos com um glide após a vogal tônica final. Camara Júnior (1991, p. 94) também cita o fato de que "pás", forma plural de "pá", e "pais", plural de "pai", são ambos pronunciados /paⁱsˈ/, com glide. O mesmo autor menciona, também, exemplos de "nus", que rima com "azuis", e "espirais", que rima com "traz". Assim, observamos que, no PB, existem registros desse fenômeno de glidificação, ou seja, inserção de um glide após a sílaba tônica final da palavra, quando esta é seguida por *–s*. Nossa primeira hipótese para os itens monossílabos favorecerem o plural de ditongo em *–u* decorre dessas considerações acerca da glidificação.

No português falado no Rio de Janeiro, os monossílabos "má", "só" e "vó" (redução de "avó") são pluralizados em [ˈmayʃ], [ˈsɔyʃ] e [ˈvɔyʃ], com a inserção dos glides após a vogal tônica. Sendo assim, os plurais de "mel", "sal" e "sol" sendo [ˈmɛyʃ], [ˈsayʃ] e [ˈsɔyʃ] (formas gramaticalmente corretas no dialeto do Rio de Janeiro) poderiam ser considerados plurais das palavras hipotéticas "*mé", "*sá" e da palavra "só", que realmente existe no PB. Para evitar essa possível correlação com os plurais de itens como "mel", "sal" e "sol", supomos que os falantes tenderam a optar pela pluralização em ditongo em *–u* para essas palavras. Uma forma de testar essa hipótese é verificar se, em dialetos que não apresentam esse tipo de glidificação, as formas de plural para "mel", "sal" e "sol" também seriam "mels", "sals" e "sols". Fizemos esse teste em Belo Horizonte, cidade que não tem como característica a inserção de glides após a vogal tônica. Elaboramos um teste sobre glidificação[86] e entrevistamos três informantes na faixa etária de 20 a 30 anos. Nesse experimento, incluímos as palavras "mel", "sal" e "sol" (os monossílabos que, em nossos experimentos, mais sofreram migração para os plurais em ditongo em *–u*) e outros itens de uma sílaba, como "pá" e "nó", que, no PB falado no Rio de Janeiro, seriam pluralizados com o glide ([ˈpayʃ] e [ˈnɔyʃ]). Nosso objetivo era verificar se os informantes entrevistados iriam pluralizar as palavras "mel", "sal" e "sol" como ditongo em *–u* e se iriam inserir glides na pluralização dos demais monossílabos (como "nó" e "pá") ou quando houvesse vogal final seguida de /s/ (como "chuchus", "nariz"). Na realização dos testes, observamos que os falantes mantiveram os plurais "méis", "sais" e "sóis". Não houve sequer um caso de migração de plurais em *–l* para ditongo em *–u*. Quanto às demais palavras propensas à glidificação, não ocorreu, também, inserção de glide em nenhum dos itens analisados, ou seja, os falantes leram, por exemplo, [kaˈʒus] e [maˈsãs], em vez de [kaˈʒuys] e [maˈsãys], como os cariocas leriam. Assim, a comparação entre dados de falantes de Belo Horizonte e do Rio de Janeiro nos faz levantar a hipótese de que o fenômeno de migração de plural em *–l* para ditongo em *–u* pode estar relacionado à glidificação na sílaba tônica final. No dialeto belo-horizontino, em que não há glidificação, não ocorreu, também, variação de plurais em itens monossílabos, pelo menos na fala dos informantes consultados para esta pesquisa.

Ainda uma última explicação possível para o fato de, em nossos dados, os monossílabos em *–l* apresentarem maior propensão ao plural em ditongo em *–u* é o fato de

[86] O "Teste de Glidificação", tal como foi apresentado aos informantes, encontra-se no "Anexo A".

que, no léxico do PB, há mais itens monossílabos em ditongo em *–u* do que em *–l*. No grupo de ditongo em *–u* no plural, o ASPA registra 29 palavras totais, das quais 13 (44,8%) são monossílabas; na classe em *–l* no plural, o ASPA apresenta 679 dados totais, dos quais apenas seis (0,88%) são monossílabos. Como há mais monossílabos em ditongo em *–u* do que em *–l*, essa parece ser, também, uma boa justificativa para que os falantes tenham aplicado o plural em ditongo em *–u* aos monossílabos em *–l*.

Há, portanto, duas hipóteses possíveis (glidificação no dialeto carioca e freqüência de monossílabos em *–l* e ditongo em *–u*) para justificar por que palavras monossílabas terminadas em *–l* favorecem a adoção do plural em ditongo em *–u*. Ambas as hipóteses parecem viáveis e apontam para o fato de que o léxico mental dos falantes comporta informações detalhadas acerca de questões fonéticas, como o fenômeno da glidificação e suas peculiaridades de pluralização. Observamos, também, que um conhecimento probabilístico acerca de padrões de freqüência lexical pode fazer com que os monossílabos em *–l* sejam líderes na adoção de plurais em ditongo em *–u*. Nos dois casos, pressupomos, de acordo com o Modelo de Redes e a Teoria dos Exemplares, representações mentais renováveis e atualizáveis a partir da experiência do falante com as palavras em seu cotidiano.

Retomando os dados da Tabela 41 acima, observamos que houve 13 casos de itens polissílabos pluralizados na regra de ditongo em *–u*. Tais itens foram: "acessível<u>s</u>", "agradável<u>s</u>", "azu<u>ls</u>", "difícil<u>s</u>" (dois casos), "faro<u>ls</u>", "funil<u>s</u>" (dois casos), "míssil<u>s</u>" (três casos) e "útil<u>s</u>" (dois casos). Essas ocorrências são pouco numerosas, visto que correspondem a 1% dos dados. Por essa razão, não faremos maiores comentários sobre elas.

O SPSS não considerou freqüência de ocorrência uma variável relevante para a migração de plurais de *–l* para ditongo em *–u*. Conforme comentamos anteriormente, nesse grupo de plurais (em *–l*), a grande significância (segundo determinada pelo SPSS) foi das variáveis número de sílabas, estrutura morfológica e vogal precedente. A fim de investigar possíveis efeitos de freqüência de ocorrência, resolvemos excluir de nossa análise essas três variáveis (número de sílabas, estrutura morfológica e vogal precedente), já que poderia estar havendo uma potencial interação entre elas. Submetemos os dados novamente ao SPSS e, nesse caso, a freqüência de ocorrência foi selecionada, com os valores apresentados na tabela abaixo:

Tabela 42: Efeito da freqüência de ocorrência nos plurais em –*l*

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	43/406	10,5	0.55
Freqüência Média	13/351	3,7	0.29
Freqüência Alta	39/355	10,9	0.65

Através da tabela acima, observamos que, se as variáveis número de sílabas, estrutura morfológica e vogal precedente forem excluídas da análise, freqüência de ocorrência é relevante. Percebe-se, então, que essas variáveis estavam, de fato, interferindo nos resultados, impedindo que os efeitos de freqüência de ocorrência fossem verificados. Retirando esses fatores, freqüência de ocorrência é selecionada, no entanto os resultados são contrários ao esperado: as palavras mais freqüentes são as líderes na migração de plural (0.65), enquanto as de freqüência baixa têm efeito quase neutro (0.55) e as de freqüência média desfavorecem o plural em ditongo em –*u* (0.29). Devemos investigar por que palavras de freqüência alta favorecem variações de plural, já que isso não era previsto em nossas hipóteses.

Observamos que, no grupo de palavras de freqüência alta, houve 39 itens terminados em –*l* que adotaram plural em ditongo em –*u*. Dessas 39 ocorrências, 35 são da palavra "gol". Sendo assim, não se pode afirmar que as palavras de alta freqüência estão liderando a adoção de plural em ditongo em –*u*, já que o item "gol" responde por quase todos esses casos. Com relação a esse item, nossa hipótese, conforme afirmamos anteriormente, é de que essa palavra entrou na língua quando o fenômeno de vocalização em –*l* já estava avançado, então ela não está em processo de variação, adotando a pluralização em ditongo em –*u*, como aconteceu com itens como "cachecols" e "gentils". Se "gol" já foi adotada pelo PB com a forma vocalizada, não pode ser considerada em conjunto com palavras que potencialmente podem vir a adotar plural em ditongo em –*u*. A fim de verificar se esse item está, realmente, interferindo nos resultados sobre freqüência de ocorrência, vamos excluí-lo momentaneamente de nossa análise. Os resultados obtidos para o SPSS, excluindo-se a palavra "gol", são:

Tabela 43: Reanálise do efeito da freqüência de ocorrência nos plurais em *–l* sem a palavra "gol"

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	43/406	10,5	0.75
Freqüência Média	13/351	3,7	0.49
Freqüência Alta	4/320	1,2	0.24

No gráfico a seguir, esses resultados podem ser visualizados de forma mais nítida:

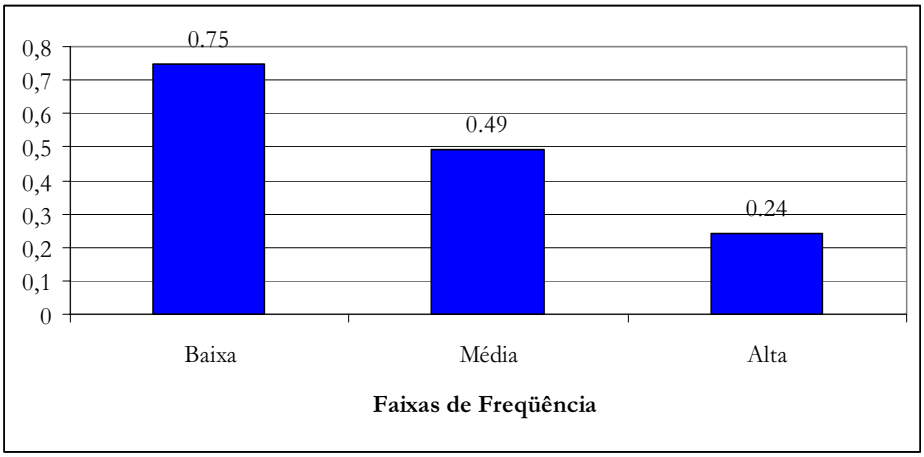


Gráfico 2: Efeito da freqüência de ocorrência nos plurais em *–l* sem a palavra "gol"

Através da tabela e do gráfico acima, observamos que, se a palavra "gol" for excluída da análise, obtemos resultados mais precisos quanto ao papel da freqüência de ocorrência: palavras de baixa freqüência favorecem a adoção de plural em ditongo em *–u* (0.75), itens de média freqüência apresentam efeito praticamente neutro (0.49) e palavras freqüentes são as menos suscetíveis a variações de plural (0.24). Esses resultados são compatíveis com nossas hipóteses de trabalho (Cf. Capítulo 4, "Quadro Teórico").

O Modelo de Redes justificaria esses efeitos de freqüência argumentando que, quando uma palavra é muito usada, sua representação mental se torna mais forte e, conseqüentemente, ela não costuma sofrer mudança analógica. Por outro lado, quando uma palavra é pouco freqüente, ela é menos acessível no léxico mental e, portanto, mais suscetível a mudanças.

Uma questão a ser ressaltada aqui é o fato de que o padrão *–l* em fim de palavra no PB é mais freqüente que o de ditongo em *–u* na mesma posição. Portanto, a

freqüência de tipo de *–l* é mais alta que a de ditongo em *–u*[87]. Então, supor-se-ia que, devido a essa interferência da freqüência de tipo, apenas palavras terminadas em ditongo em *–u* migrassem para a classe de *–l*. No entanto, nossos dados mostram que o contrário (palavras terminadas em *–l* adotarem pluralização em ditongo em *–u*) também ocorre. Provavelmente, a falta de distinção fonética entre *–l* e *–u* em fim de palavra fez com que a referência para pluralização (`[l]` ou `[w]`) fosse perdida, portanto *–is* e *–us* são ambos válidos para pluralizar palavras que terminem com o segmento `[w]`. No entanto, em termos de freqüência de tipo, ainda é possível que algum efeito seja encontrado, no sentido de palavras terminadas em ditongo em *–u* sofrerem mais variação de plural do que palavras terminadas em *–l*. Isso será analisado na próxima subseção, quando apresentarmos os resultados para as palavras terminadas em ditongo em *–u*.

O último fator considerado relevante pelo programa SPSS foi **escolaridade**, cujas possibilidades eram Ensino Fundamental ou Superior. Vejamos os resultados abaixo:

Tabela 44: Efeito da escolaridade nos plurais em *–l*

Nível de Escolaridade	**N**	**%**	**P.R.**
Ensino Fundamental	63/545	11,5	0.71
Ensino Superior	32/567	5,6	0.28

Os falantes de Ensino Fundamental demonstram favorecimento do plural em ditongo em *–u* para itens terminados em *–l* (peso relativo de 0.71), ao passo que os falantes de Nível Superior tendem a evitar o fenômeno (peso relativo de 0.28). Nossas hipóteses para justificar esse resultado são duas: 1) Primeiro, o falante com mais anos de escolaridade pode ser, em geral, mais atento a questões relacionadas à linguagem. Conseqüentemente, tal indivíduo teria um cuidado maior com a fala e evitaria formas lingüísticas consideradas não-padrão pela gramática normativa; 2) Em segundo lugar, o nível mais alto de escolarização pode levar os falantes a um contato maior com a língua escrita, modalidade essa em que há mais preocupação gramatical e as formas padronizadas prevalecem. A escrita, em geral, preserva as formas de plural padrão, por isso apresenta menos palavras terminadas em *–l* pluralizadas como ditongo em *–u*. Talvez por isso os falantes com nível de escolaridade

[87] Exceção deve ser feita quanto aos monossílabos em *–l* e ditongo em *–u*. Conforme vimos nas páginas anteriores, os monossílabos em ditongo em *–u* são em maior quantidade que os em *–l*. Portanto, embora a freqüência geral de ocorrência para itens em *–l* seja maior que a de itens em ditongo em *–u*, na classe dos monossílabos essa situação se inverte.

mais alto (e, conseqüentemente, com maior acesso à língua escrita) tenham evitado adotar plural em ditongo em *–u* para itens terminados em *–l*. O falante menos escolarizado, por sua vez, teria menos acesso à língua escrita, por isso pode ser que seu léxico mental contenha mais formas generalizadas, já que, normalmente, é na modalidade falada que se originam as variações lingüísticas.

Depois de apresentarmos os resultados relevantes para os plurais em *–l*, passaremos à subseção seguinte, em que analisamos os itens terminados em ditongo em *–u*.

6.3 Resultados para os plurais terminados em ditongo em *–u*

Para o grupo de plurais em *–u*, coletamos um total de 971 dados. A variável dependente era estratificada da seguinte maneira: 1) Plural em ditongo em *–u*; 2) Plural em *–l*; 3) Ausência de plural; 4) Outro plural[88]. Os resultados gerais são apresentados na tabela abaixo:

Tabela 45: Resultados gerais para os plurais em ditongo em *–u*

Formas de Plural	N	%
Plural em ditongo em *–u*	762/971	78,4
Plural em *–l*	192/971	19,7
Ausência de plural	17/971	1,7

Através da tabela acima, observamos que o número de itens em ditongo em *–u* pluralizados como *–l* é de 192, ou seja, 19,7% dos dados. Esse número é maior que os itens em *–l* que foram pluralizados em *–u* (95 dados ou 8,1%, conforme a Tabela 34). Observamos, então, um efeito da freqüência de tipo, porque, conforme mencionamos no Capítulo 3, "Revisão de Literatura", no léxico do PB, o grupo de palavras terminadas em *–l* é maior do que o de palavras terminadas em *–u* (679 dados de *–l* no plural contra 29 de ditongo em *–u*, também no plural, segundo o ASPA[89]). Sendo assim, é de se esperar que

[88] Nos plurais em ditongo em *–u*, não houve casos de utilização de outra forma de plural, diferente da forma em *–l* ou em ditongo em *–u*. Portanto, não incluímos esse caso na Tabela 45.

[89] Conforme vimos nas Tabelas 10 e 11, o Novo Dicionário Aurélio Eletrônico também ratifica a maior freqüência de tipo da classe em *–l* sobre a de ditongo em *–u*.

essa freqüência dos itens em *–l* acabe por "atrair" os membros da classe de ditongo em *–u*, já que a distinção fonética entre *–l* e *–u* no fim de sílaba foi perdida (Cf. CRISTÓFARO-SILVA, 2002, p. 63). Os resultados para os plurais de ditongo em *–u* comprovam, portanto, os pressupostos do Modelo de Redes, que argumenta que é a freqüência de tipo a responsável pela produtividade de uma classe paradigmática e, em termos de mudanças analógicas, palavras de tipos menos freqüentes tendem a migrar para classes mais freqüentes, sobretudo os itens que, individualmente, apresentarem baixa freqüência de ocorrência.

Os 17 casos de ausência de plural foram: "ateu", "berimbau" (seis casos), "chapéu" (três casos), "degrau", "grau" (dois casos), "mau", "pneu" e "troféu" (dois casos). A palavra que mais sofreu apagamento de plural foi "berimbau", que está na faixa de freqüência baixa do Corpus NILC/São Carlos. Os casos de ausência de plural ocorreram, sobretudo, na etapa do "Conjunto de Figuras". Em situações em que o falante via vários berimbaus juntos, a frase pronunciada era, por exemplo, "Um conjunto de berimbau" ou, na figura que mostrava um móvel repleto de chapéus, o falante pronunciava, por exemplo, "Um cabide cheio de chapéu". Nessas ocorrências, observamos a utilização de uma palavra no singular para referir-se a um conjunto de itens no plural. Não mencionaremos mais informações sobre essa peculiaridade do PB, pois tal assunto já foi abordado na subseção 6.1, quando tratamos dos resultados gerais para os plurais em *–ão*.

Nos parágrafos subseqüentes, analisaremos os resultados para as variáveis segundo as quais nossos dados foram codificados. Nessa análise, excluiremos os dados relativos à ausência de plural, pois já mencionamos esses casos no parágrafo anterior. Após essas exclusões, contamos com um total de 954 dados em que não houve apagamento das marcas formais de plurais. Submetemos esses dados ao programa SPSS e, na regressão binária, o programa considerou relevantes as variáveis **palavra**, **faixa etária**, **escolaridade** e **gênero**. Discutiremos, primeiramente, os resultados para **palavra**, cuja tabela é apresentada abaixo[90]:

[90] Os resultados numéricos, percentuais e de peso relativo referem-se à aplicação do plural em *–l* para palavras terminadas em ditongo em *–u*. Maiores esclarecimentos sobre isso já foram feitos nos resultados dos plurais anteriores (Cf. Subseções 6.1 e 6.2 deste capítulo).

Tabela 46: Efeito da palavra nos plurais em ditongo em *–u*

Palavras	Freqüência	N	%	P.R.
1. Berimbau	baixa – 0	27/62	43,5	0.99
2. Jirau	baixa – 01	15/35	42,8	0.99
3. Mausoléu	baixa – 04	13/35	37,1	0.99
4. Ateu	**baixa – 24**	**0/34**	**0**	**0**
5. Véu	baixa – 33	9/36	25	0.99
6. Pau	**baixa – 46**	**0/36**	**0**	**0**
7. Troféu	baixa – 72	48/101	47,5	0.99
8. Degrau	média – 109	21/82	25,6	0.99
9. Chapéu	média – 125	43/102	42,1	0.99
10. Céu	média – 163	1/36	2,7	0.88
11. Réu	média – 309	8/36	22,2	0.99
12. Mau	média – 346	1/35	2,8	0.88
13. Museu	média – 424	1/36	2,7	0.88
14. Pneu	alta – 554	1/68	1,4	0.78
15. Judeu	alta – 808	2/35	5,7	0.94
16. Grau	alta – 916	1/77	1,2	0.75
17. Europeu	alta – 1.119	1/36	2,7	0.88
18. Meu	**alta – 2.139**	**0/36**	**0**	**0**
19. Seu	**alta – 29.682**	**0/36**	**0**	**0**

Na tabela acima, estão destacados em negrito todos os itens que não sofreram nenhuma migração de plurais de ditongo em *–u* para *–l*. Essas palavras são: "ateu", "pau", "meu" e "seu". Dentre essas palavras, "ateu" e "pau" são de baixa freqüência de ocorrência e "meu" e "seu" são de alta freqüência de ocorrência. Nesse caso, não podemos afirmar que houve efeito de freqüência de ocorrência nos itens léxicos individualmente, já que palavras de freqüências diferentes foram mais conservadoras no fenômeno.

Analisando os itens da tabela acima, observamos, novamente, que é necessário fazer uma seleção em nosso banco de dados sobre plurais em ditongo em *–u*. Os pesos relativos categóricos podem estar mascarando possíveis resultados probabilísticos para migrações de plurais nessa classe. Excluímos, então, todos os casos em que não houve nenhuma adoção de plural em *–l*. Portanto, foram removidas as ocorrências das palavras "ateu", "pau", "meu" e "seu". Ficamos, nesse caso, com 812 dados, cujos resultados são apresentados na tabela a seguir:

Tabela 47: Reanálise do efeito da palavra nos plurais em ditongo em *–u*

Palavras	Freqüência	N	%	P.R.
1. Berimbau	**0 – baixa**	**27/62**	**43,5**	**0.93**
2. Jirau	**01 – baixa**	**15/35**	**42,8**	**0.92**
3. Mausoléu	04 – baixa	13/35	37,1	0.88
4. Véu	33 – baixa	9/36	25	0.73
5. Troféu	**72 – baixa**	**48/101**	**47,5**	**0.94**
6. Degrau	109 – média	21/82	25,6	0.76
7. Chapéu	**125 – média**	**43/102**	**42,1**	**0.92**
8. Céu	163 – média	1/36	2,7	0.10
9. Réu	309 – média	8/36	22,2	0.68
10. Mau	346 – média	1/35	2,8	0.11
11. Museu	424 – média	1/36	2,7	0.10
12. Pneu	554 – alta	1/68	1,4	0.05
13. Judeu	808 – alta	2/35	5,7	0.21
14. Grau	916 – alta	1/77	1,2	0.04
15. Europeu	1.119 – alta	1/36	2,7	0.10

Excluindo os dados em que não houve nenhum caso de migração de plural, obtivemos resultados mais adequados sobre as palavras favorecedoras dos plurais em *–l*. Em negrito destacamos os itens líderes na migração de plural em ditongo em *–u* em direção a *–l*: "berimbau", "jirau", "troféu" e "chapéu". As palavras "berimbau", "jirau" e "troféu" pertencem à freqüência baixa de ocorrência, ao passo que o item "chapéu" tem freqüência média. Observamos, portanto, que não houve nenhum item de freqüência de ocorrência alta liderando o processo de migração de plural do grupo em ditongo em *–u* para *–l*. Concluímos, nesse caso, que a freqüência da palavra individualmente é um indício para que ela mantenha sua classe etimológica ou migre para um grupo que apresente um padrão analógico mais freqüente.

Feitos os comentários sobre a palavra individualmente, retiramos essa variável de nossa análise, a fim de verificar a interferência de outros fatores nas variações dos plurais em ditongo em *–u*. Submetendo novamente os dados ao programa SPSS (sem o fator palavra), as variáveis lingüísticas consideradas relevantes foram **vogal precedente**, **estrutura morfológica**, **número de sílabas** e **freqüência de ocorrência**; as variáveis

extralingüísticas selecionadas foram **faixa etária**, **escolaridade** e **gênero**[91]. Comentaremos cada um desses fatores separadamente.

Primeiramente, apresentamos os resultados para **vogal precedente**. As possibilidades eram vogal precedente [a], como "berimbau", "grau", vogal precedente [ɛ], como "céu", "mausoléu", e vogal precedente [e], como "museu", "pneu"[92]. Vejamos os resultados abaixo:

Tabela 48: Efeito da vogal precedente nos plurais em ditongo em *–u*

Vogal Precedente	N	%	P.R.
[a]	65/327	19,8	0.71
[ɛ]	122/346	35,2	0.83
[e]	5/281	1,7	0.07

Na tabela acima, observa-se que as vogais abertas ([a] e [ɛ]) favorecem a migração de plurais de ditongo em *–u* para *–l* e a vogal fechada ([e]) desfavorece. Os itens terminados em [ɛw] pluralizados como [ɛys] foram: "céu" (um caso), "chapéu" (43 casos), "mausoléu" (13 casos), "réu" (oito casos), "troféu" (48 casos) e "véu" (nove casos). Houve apenas cinco casos de palavras terminadas em [ew] que adotaram plural em *–l*: "europeu", "judeu" (dois casos), "museu" e "pneu". Essas palavras todas são de baixa e média freqüência de ocorrência. No grupo de palavras em [aw] (que também favoreceu o plural em *–l*), as líderes nas migrações de plurais foram "berimbau", "degrau" e "jirau", também de baixa e média freqüência de ocorrência. Nossa hipótese, portanto, é de que o suposto favorecimento dos segmentos [aw] e [ɛw] decorre, na verdade, de efeitos de freqüência, que fazem com que tais palavras desses grupos, por serem pouco usadas, adotem plural diferente daquele que sua classe tem. Esses efeitos serão discutidos posteriormente, no desenrolar desta análise.

Quanto ao fator **estrutura morfológica**, estávamos analisando se a palavra apresentava sufixo, como "europeu", "judeu", ou não, como "berimbau", "museu". Vejamos os resultados na tabela abaixo:

[91] Na análise dos plurais em ditongo em *–u*, o SPSS não excluiu, portanto, nenhuma das variáveis lingüísticas ou extralingüísticas codificadas para esse grupo de palavras. Houve relevância de todos os fatores.

[92] Conforme afirmamos anteriormente, os itens em *–iu*, *–ou* e *–uu* não foram considerados em nossa análise porque representavam um baixo quantitativo de dados (Cf. Tabelas 9 e 11).

Tabela 49: Efeito da estrutura morfológica nos plurais em ditongo em *–u*

Estrutura Morfológica	N	%	P.R.[93]
Sufixo	3/71	4,2	0.27
Não-sufixo	189/883	21,4	0.72

Os números da tabela acima demonstram que, quando a palavra tem sufixo, tende a manter o plural em ditongo em *–u* (0.27) e, quando não tem sufixo, favorece plural em *–l* (0.72). Esse resultado é análogo àquele encontrado para os plurais em *–l*, já que, naquele grupo, os itens sem sufixo também foram os líderes na adoção de plurais em ditongo em *–u*. Conforme afirmamos anteriormente, a classe em *–u* apresenta apenas dois sufixos: *–eu* e *–aréu*. Nossa hipótese para a manutenção do plural em ditongo em *–u* no caso de palavra com sufixo é a seguinte: como o grupo terminado em ditongo em *–u* tem apenas dois sufixos, essa classe (de itens com sufixo) acaba sendo marcada, torna-se mais forte no léxico mental e, conseqüentemente, é menos suscetível a generalizações em direção ao padrão mais freqüente (em *–l*). A partir das similaridades semânticas e fonológicas entre os itens em *–u* estocados no léxico mental, a informação morfológica sobre a existência de sufixo emerge, prevenindo tais palavras (com sufixo) contra eventuais efeitos analógicos. Por outro lado, quando a palavra não tem sufixo, estabelece conexões menos coesas no léxico mental e, por isso, torna-se mais propensa a adotar o plural em *–l*.

A seguir, apresentamos os dados sobre **número de sílabas**, cujas alternativas eram: monossílabos, como "grau", "véu", ou polissílabos, como "chapéu", "europeu". Os resultados podem ser visualizados na tabela abaixo:

Tabela 50: Efeito do número de sílabas nos plurais em ditongo em *–u*

Número de Sílabas	N	%	P.R.
Monossílabos	20/328	6	0.25
Polissílabos	172/626	27,4	0.74

A tabela acima mostra que os monossílabos desfavorecem o plural em *–u* (peso relativo de 0.25) e os polissílabos favorecem (peso relativo de 0.74). Os 20 casos de

[93] Para que os resultados de peso relativo fossem compatíveis com as porcentagens apresentadas, foi necessário excluir as variáveis vogal precedente, freqüência e número de sílabas, pois elas estavam interagindo com estrutura morfológica, fazendo com que os pesos relativos fossem inadequados para os valores percentuais.

monossílabos terminados em ditongo em *–u* que adotaram pluralização em *–l* foram: "céu", "grau", "mau", "réu" (oito vezes) e "véu" (nove vezes). Observamos que "réu" e "véu", os dois monossílabos líderes na adoção do plural em *–l* para itens em ditongo em *–u*, são palavras de baixa freqüência. Supomos, então, que, em geral, os monossílabos são menos suscetíveis a migrações de plural, mas, se sua freqüência de ocorrência for baixa, é possível que tais itens também adotem plural em *–l*, por causa da alta freqüência de tipo desse paradigma.

Na análise dos plurais em *–l*, número de sílabas também foi uma variável relevante. Naquele grupo, observamos que os monossílabos favoreceram a adoção de plurais em ditongo em *–u* (0.96), ao passo que os polissílabos desfavoreceram (0.03) (Cf. Tabela 41). Existe, portanto, uma coerência nos resultados sobre número de sílabas para plurais em *–l* e ditongo em *–u*: a classe de itens em *–l* é maior que a de ditongo em *–u*, no entanto, como vimos anteriormente, há mais monossílabos no grupo de ditongo em *–u* que no de *–l*. Por causa disso, os falantes, quando têm de pluralizar uma palavra terminada em [w] (com *–l* ou *–u* ortográficos), tendem a adotar plural em *–l* para itens polissílabos e plural em ditongo em *–u* para palavras monossílabas. É isso o que os resultados demonstram, já que, no grupo em *–l*, monossílabos favoreceram plural em *–u* e, no grupo em ditongo em *–u*, polissílabos favoreceram plural em *–l*. Assim, parece que, no léxico mental, há informações probabilísticas sobre as duas classes de plurais (em *–l* e ditongo em *–u*) e a pluralização quanto ao número de sílabas é uma generalização que deriva das conexões entre os itens listados no léxico.

Os próximos resultados são para **freqüência de ocorrência**. As faixas determinadas foram: freqüência baixa (de zero até 99 ocorrências), média (de 100 a 500) e alta (mais de 500 ocorrências). Apresentamos, abaixo, a tabela com os resultados:

Tabela 51: Efeito da freqüência de ocorrência nos plurais em ditongo em *–u*

Faixas de Freqüência	**N**	**%**	**P.R.**
Freqüência Baixa	112/339	33	0.67
Freqüência Média	75/327	22,9	0.56
Freqüência Alta	5/288	1,7	0.27

O gráfico abaixo facilita a visualização dos resultados:

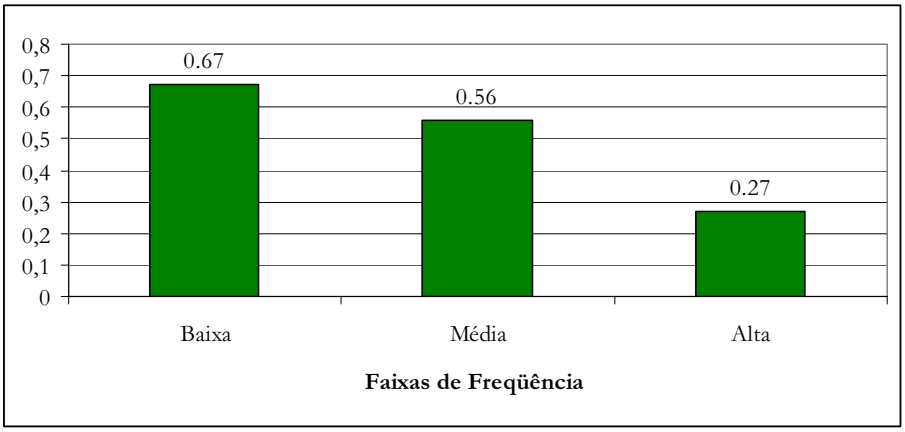


Gráfico 3: Efeito da freqüência de ocorrência nos plurais em ditongo em *–u*

Os resultados para freqüência de ocorrência são coerentes com nossa hipótese de trabalho (segundo o Modelo de Redes): palavras pouco freqüentes favorecem a adoção do plural em *–l* (peso relativo de 0.67), ao passo que palavras freqüentes são mais fortes no léxico mental e desfavorecem os plurais analógicos (peso relativo de 0.27). Os cinco itens de alta freqüência que migraram para a classe de *–l* foram: "europeis", "grais", "judeis" (dois casos) e "pneis".

Mais uma vez, os dados demonstram interação entre as freqüências de tipo e de ocorrência, da seguinte maneira: a alta freqüência de tipo faz com que um padrão na língua se torne mais forte e, conseqüentemente, mais produtivo. Se uma palavra pertencente a um padrão pouco freqüente na língua (como os plurais em ditongo em *–u*) apresenta, também, baixa freqüência de ocorrência, ela se torna mais suscetível a migrar para o padrão mais freqüente na língua, no caso o de plurais em *–l*. Assim, observamos que, dentro de um paradigma lingüístico, em geral são os itens com baixa freqüência de ocorrência os primeiros a migrarem para o padrão mais recorrente. Para que uma palavra pertencente a uma classe de baixa freqüência de tipo (como a de plurais em ditongo em *–u*) mantenha seu plural etimológico, ela precisa ter alta freqüência de ocorrência. Se a palavra pertence a um padrão pouco freqüente e, além disso, apresenta baixa freqüência de ocorrência, é provável que ela acabe sendo atraída para um padrão mais freqüente na língua e com o qual ela mantenha similaridades fonológicas, como é o caso da migração dos plurais em ditongo em *–u* para a classe em *–l*.

Segundo a Teoria de Palavras e Regras, não deveria haver efeitos de freqüência em itens regulares, mas não é isso o que se observa no grupo de palavras terminadas em ditongo em *–u*. Esses itens são pluralizados através da adição do morfema *–s*, o que

caracteriza o padrão regular e mais freqüente de plurais no PB. No entanto, em vez de simplesmente adotarem o plural em *–s*, os falantes, em alguns casos, utilizaram o plural em *–is*, produzindo uma mudança de radical, em vez de flexão. As similaridades fonológicas entre as classes em *–l* e ditongo em *–u* permitem que haja migrações de plurais de um paradigma pouco freqüente (ditongo em *–u*) para outro mais freqüente (em *–l*). Como, no singular, a distinção fonética entre os dois grupos foi perdida (ambas terminam em [w], a despeito da ortografia em *–l* ou*–u*), existe um ambiente fonético propício para que generalizações baseadas em freqüência afetem essas classes. Observa-se, novamente, que o léxico mental comporta informações probabilísticas em termos da freqüência de tipos e de ocorrências para padrões lingüísticos e palavras individualmente.

Os três fatores extralingüísticos selecionados pelo programa SPSS serão comentados nos parágrafos subseqüentes. Primeiramente, mostramos os resultados para a **faixa etária**, cuja distribuição era: jovens (de 15 a 20 anos), medianos (de 35 a 40 anos) e adultos (de 55 a 60 anos). Vejamos os resultados abaixo:

Tabela 52: Efeito da faixa etária nos plurais em ditongo em *–u*

Faixa Etária	N	%	P.R.
Jovens	106/314	33,7	0.77
Medianos	54/329	16,4	0.42
Adultos	32/311	10,2	0.28

Observamos que, para os plurais em ditongo em *–u*, os números indicam mudança em progresso, no sentido de que os falantes mais jovens favorecem os plurais em *–l* (peso relativo de 0.77), ao passo que os falantes mais idosos demonstram tendência à preservação dos plurais etimológicos (peso relativo de 0.28 para adoção de plurais em *–l*). Como a Teoria dos Exemplares afirma, falantes mais velhos tendem a ter mais exemplares para a mesma categoria do que falantes mais jovens; por outro lado, o léxico mental dos falantes mais velhos não é atualizado tão rapidamente quanto o dos falantes jovens, porque a memória decai com a idade (PIERREHUMBERT, 2001b, p. 149). Tais fatos parecem justificar os resultados dos plurais em ditongo em *–u* no que diz respeito à faixa etária.

Os próximos resultados são para níveis de **escolaridade** (Fundamental ou Superior):

Tabela 53: Efeito da escolaridade nos plurais em ditongo em *–u*

Nível de Escolaridade	N	%	P.R.
Ensino Fundamental	123/478	25,7	0.63
Ensino Superior	69/476	14,4	0.36

Falantes de Ensino Fundamental favorecem as generalizações de plural (peso relativo de 0.63) e informantes com Nível Superior as desfavorecem (peso relativo de 0.36). À semelhança do que ocorreu com os plurais em *–l* (Cf. Tabela 44), parece que a escola exerce papel importante no sentido de manter as formas consideradas corretas, de acordo com a gramática padrão da língua. Cristófaro-Silva, Gomes, Guimarães e Huback (2005), analisando crianças na faixa etária de três a 12 anos, também observaram que, quanto maior o nível de escolaridade, mais os plurais etimológicos são formalmente marcados.

Finalmente, os resultados para **gênero** (feminino ou masculino), último fator selecionado pelo programa SPSS para os plurais de ditongo em *–u*, se encontram logo abaixo:

Tabela 54: Efeito do gênero nos plurais em ditongo em *–u*

Gênero	N	%	P.R.
Feminino	73/481	15,1	0.38
Masculino	119/473	25,1	0.61

Com relação ao gênero, observamos que os homens favorecem a utilização de plurais em *–l* para palavras terminadas em ditongo em *–u* (peso relativo de 0.61) e as mulheres desfavorecem o mesmo fenômeno (peso relativo de 0.38). Nossa hipótese para isso é o fato de que, apesar de o léxico mental dos falantes ser constantemente atualizado, as mulheres, segundo a literatura lingüística comenta, são mais cuidadosas com a linguagem, tendem a evitar variantes não-padrão e demonstram ter mais consciência do prestígio que a fala possui (Cf. CHAMBERS, 1995). Sendo assim, apesar de estocarem as formas generalizadas, as mulheres tenderiam a evitar sua utilização, talvez por causa do estigma que recai sobre essas formas. Não fizemos teste de reação subjetiva com os informantes, a fim de verificar se sobre formas como "chapéis" ou "troféis", por exemplo, incide algum preconceito lingüístico. No entanto, parece pertinente sugerir que as mulheres

tendam a preservar os plurais corretos, de acordo com a norma culta, justamente porque são mais sensíveis a formas de prestígio (Cf. LABOV, 1972).

Finalizados os comentários gerais sobre os dados coletados, passaremos para a próxima subseção, que apresenta os resultados alcançados com o "Teste de Reação" para os grupos de plurais analisados.

6.4 Resultados para o "Teste de Reação"

Nesta subseção, analisamos os dados coletados através do "Teste de Reação". Avaliávamos, neste experimento, não apenas as formas que os informantes adotaram como plurais para as palavras apresentadas, mas também a maneira como os falantes reagiram quando proveram as respostas. Conforme explicado na subseção 5.1.2 do Capítulo 5, "Metodologia", estávamos considerando se houve ou não hesitação dos falantes no momento em que pronunciaram as respostas para as palavras apresentadas pela pesquisadora. Nesse teste, portanto, a variável dependente era "hesitação/não-hesitação" por parte do falante[94]. No caso de haver hesitação, investigamos quais eram as variáveis relevantes para que isso ocorresse. Discutiremos cada um dos grupos de plurais em separado.

Com relação aos plurais em *–ão*[95], contabilizamos 641 dados extraídos no "Teste de Reação"[96]. Ao todo, houve 76 hesitações, o que corresponde a 11,8% do total dos dados. Submetendo esses dados ao SPSS, foram selecionados os fatores **palavra** e **número de sílabas**. As hesitações por **palavra** encontram-se na tabela abaixo:

---

[94] Em todas as tabelas, os números, porcentagens e pesos relativos devem ser considerados em relação à existência de hesitação, já que essa foi a variante a partir da qual os pesos relativos foram calculados.

[95] No "Teste de Reação", foram mantidos os plurais etimológicos em *–ões*, ao contrário do que fizemos na análise geral dos dados para as palavras terminadas em *–ão* no singular. Consideramos que, nesse teste, a variável relevante a ser analisada era hesitação ou não-hesitação do falante no momento de pronunciar o plural para a palavra dada. Já que não estávamos investigando migrações de plural em uma determinada direção, o fato de a palavra ser pluralizada etimologicamente em *–ãos*, *–ães* ou *–ões* não interferia nos resultados para hesitações.

[96] No grupo de *–ão*, só havia um caso de apagamento de plural, por isso esse dado foi excluído da análise do "Teste de Reação".

Tabela 55: Efeito da palavra no "Teste de Reação" para plurais em *–ão*

Palavras	Freqüência	N	%	P.R.
1. Limão	**baixa – 13**	**7/36**	**19,4**	**0.99**
2. Escrivão	baixa – 14	6/35	16,6	0.99
3. Bênção	**baixa – 26**	**9/35**	**25,7**	**0.99**
4. Guardião	baixa – 26	6/35	16,6	0.99
5. União	**baixa – 34**	**10/35**	**28,5**	**0.99**
6. Vulcão	*baixa – 62*	*0/36*	*0*	*0*
7. Pão	*média – 118*	*0/36*	*0*	*0*
8. Anão	média – 142	4/36	11,1	0.99
9. Religião	*média – 191*	*0/36*	*0*	*0*
10. Cristão	**média – 235**	**7/35**	**20**	**0.99**
11. Cão	média – 350	4/35	11,4	0.99
12. Exposição	média – 494	4/36	11,1	0.99
13. Opinião	alta – 571	1/36	2,7	0.99
14. Avião	alta – 874	1/36	2,7	0.99
15. Alemão	alta – 1.025	4/36	11,1	0.99
16. Irmão	alta – 1.125	4/35	11,4	0
17. Órgão	alta – 1.578	5/36	13,8	0.99
18. Mão	alta – 2.949	4/36	11,1	0

Na tabela acima, observamos que os pesos relativos quanto às hesitações nos plurais em *–ão* são categóricos, variando entre zero e 0.99. A fim de obter pesos relativos mais graduais, removemos os itens de peso relativo categórico em zero, ou seja, "vulcão", "pão", "religião", "irmão" e "mão". Ficamos, então, com 462 dados (13 palavras) e, submetendo-os ao SPSS, a variável palavra deixa de ser selecionada, por isso não pudemos obter resultados de peso relativo para cada um dos itens individualmente. No entanto, analisando os valores percentuais apresentados na tabela acima, marcamos em negrito os itens mais suscetíveis a hesitações e em itálico os que não sofreram hesitações. Observamos que as palavras que mais provocaram dúvidas nos falantes foram "união", "bênção", "cristão" e "limão" (nessa ordem). Dessas quatro palavras, "união", "bênção" e "limão" são de freqüência de ocorrência baixa e "cristão", de média. Por outro lado, houve três itens ("pão", "religião" e "vulcão") que não causaram nenhuma hesitação nos falantes. Essas palavras são de freqüência baixa e média de ocorrência. Parece, então, estar havendo uma interferência do fator freqüência de ocorrência no "Teste de Reação". Esse ponto será comentado de forma mais aprofundada ainda no decorrer desta análise.

A exemplo do que fizemos com os dados gerais sobre plurais em *–ão*, excluímos a variável palavra, porque sua estratificação era grande demais (18 itens) e isso faria com que esse fator fosse, inevitavelmente, considerado relevante pelo SPSS; além disso, conforme já mencionamos anteriormente, a interação da variável palavra com outros fatores lingüísticos interferiria nos resultados gerais. Como já fizemos os comentários sobre cada item individualmente, excluímos a variável palavra e submetemos novamente os dados ao SPSS. O programa apontou relevância dos fatores **tipo de plural** (resposta adotada pelo falante) e **freqüência de ocorrência**. Com relação à variável **tipo de plural**, os resultados são os seguintes:

Tabela 56: Efeito do tipo de plural no "Teste de Reação" para plurais em *–ão*

Tipo de Plural (Resposta do informante)	**N**	**%**	**P.R.**
Plural em *–ões*	33/323	10,2	0.43
Plural em *–ãos*	31/175	17,7	0.65
Plural em *–ães*	12/143	8,3	0.41

Observamos que os plurais em *–ãos* favorecem as hesitações (peso relativo de 0.65), ao passo que os plurais em *–ães* e *–ões* as desfavorecem (pesos relativos de 0.41 e 0.43, respectivamente). Talvez isso tenha ocorrido porque, nesse grupo em *–ão*, a maior corrente de migração de plurais é de *–ãos* para *–ões* (Cf. Tabela 25). No caso de *–ães* migrando para *–ões*, houve alguns poucos casos isolados, conforme já discutimos anteriormente, na mesma tabela. Então, nossa hipótese para justificar por que itens em *–ãos* causam maiores hesitações é o fato de que, quando vão pronunciar esses plurais, os falantes ficam em dúvida se a forma adequada é *–ãos* ou *–ões*, por isso hesitam enquanto buscam, em seu léxico mental, o melhor exemplar para as palavras. As migrações de *–ãos* para *–ões* são mais comuns, então é possível que haja formas em competição no léxico mental dos falantes quando se trata de pluralizar palavras etimológicas em *–ãos*. Com relação ao plural em *–ães*, já vimos que esse é um grupo pequeno no PB (12 itens, segundo o ASPA[97]), composto por itens monossílabos que, em geral, não adotam plurais em *–ões*. No caso de *–ões*, ocorre desfavorecimento das hesitações (peso relativo de 0.43) porque esse já é o padrão mais freqüente para a classe terminada em *–ão* no singular, por isso sua alta

[97] O Novo Dicionário Aurélio Eletrônico também registra que o morfema *–ães* é o menos freqüente para pluralizar itens em *–ão* no singular (Cf. Tabela 6).

freqüência de tipo faria com que sua adoção para a pluralização sofresse menos restrições relacionadas à competição de exemplares. No caso de haver mais de um exemplar no léxico mental para a mesma palavra, o plural em *–ões* tende a vencer, já que tem maior freqüência de tipo. Por isso, palavras pluralizadas em *–ões* causam menos hesitações nos falantes.

A seguir, analisaremos os resultados sobre **freqüência de ocorrência** para as hesitações nos plurais em *–ão*:

Tabela 57: Efeito da freqüência de ocorrência no "Teste de Reação" para plurais em *–ão*

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	38/213	17,8	0.64
Freqüência Média	20/214	9,3	0.49
Freqüência Alta	18/214	8,4	0.36

A variável freqüência de ocorrência apresenta resultados semelhantes aos que já haviam sido discutidos anteriormente, quando analisamos os dados de plurais em *–ão* globalmente (Cf. Tabela 31). Mais uma vez, observamos que as palavras menos freqüentes favorecem as hesitações de plural (peso relativo de 0.64), palavras de média freqüência têm efeito neutro (peso relativo de 0.49) e palavras mais freqüentes inibem as hesitações nas respostas (peso relativo de 0.36 para freqüência alta). Esse resultado comprova o que o Modelo de Redes e o de Exemplares postulam: palavras mais freqüentes têm representações mais fortes no léxico mental e, por isso, são acessadas mais facilmente. Por outro lado, palavras de baixa freqüência de ocorrência são menos disponíveis no léxico mental e, conseqüentemente, são acessadas de forma mais lenta pelos falantes.

Quanto aos plurais em *–l*, houve 646 dados[98] para o "Teste de Reação". As hesitações somaram 98 ocorrências, ou seja, 15,1% do total de dados. Na primeira rodada do SPSS, foram selecionadas as variáveis **palavra** e **número de sílabas**. Quanto à variável **palavra**, os resultados são apresentados a seguir:

[98] No Teste de Reação para plurais em *–l* e ditongo em *–u*, foram mantidas as variáveis "Outro plural" e "Ausência de plural" (quando esses casos ocorreram). Consideramos que o mais relevante era a reação do informante (com ou sem hesitação) à palavra ouvida, independentemente de o plural ter sido apagado ou de ter sido usada uma forma não-padrão. No caso dos plurais em *–ão*, houve um caso de apagamento de plural e nenhum caso de outro plural, por isso excluímos de nossa análise esse dado de ausência de plural. Nos plurais em *–l* e ditongo em *–u*, como essas ocorrências (de ausência de plural e outro plural) são mais numerosas (11 para *–l* e três para ditongo em *–u*), decidimos mantê-las em nossa análise.

Tabela 58: Efeito da palavra no "Teste de Reação" para plurais em *–l*

Palavras	Freqüência	N	%	P.R.
1. Mel	**baixa – 0**	**18/36**	**50**	**0.99**
2. Anzol	*baixa – 08*	*0/36*	*0*	*0*
3. Avental	baixa – 17	4/36	11,1	0.98
4. Sol	**baixa – 27**	**17/36**	**47,2**	**0.99**
5. Gentil	baixa – 31	1/36	2,7	0.92
6. Agradável	*baixa – 74*	*0/36*	*0*	*0*
7. Anel	*média – 110*	*0/35*	*0*	*0*
8. Lençol	média – 118	1/36	2,7	0.92
9. Sal	**média – 129**	**16/36**	**44,4**	**0.99**
10. Acessível	média – 165	5/36	13,8	0.98
11. Azul	média – 280	4/36	11,1	0.98
12. Infantil	média – 328	2/36	5,5	0.96
13. Espanhol	alta – 501	5/35	14,2	0.98
14. Difícil	alta – 654	9/36	25	0.99
15. Útil	alta – 783	6/36	16,6	0.98
16. Hospital	alta – 1.513	1/36	2,7	0.92
17. Responsável	*alta – 1. 894*	*0/36*	*0*	*0*
18. Gol	alta – 4.565	9/36	25	0.99

Na tabela acima, novamente observamos que os pesos relativos não apresentam gradualidade compatível com os valores percentuais. Mais uma vez, percebemos que é necessário remover as palavras de peso relativo categórico, ou seja, aqueles itens que não causaram nenhuma hesitação nos falantes. Essas palavras são "anzol", "agradável", "anel" e "responsável". Retirando esses itens, ficamos com 503 dados (14 palavras). Submetendo-os ao SPSS, a variável palavra deixa de ser selecionada. No entanto, da mesma forma como procedemos para o "Teste de Reação" sobre plurais em *–ão*, marcamos, na tabela acima, em negrito os itens favorecedores de hesitações e, em itálico, os itens que inibiram hesitações (considerando as distribuições percentuais). Os itens que mais causaram dúvidas nos falantes foram "mel", "sol" e "sal" (nessa ordem). Os dois primeiros são de baixa freqüência de ocorrência e "sal" pertence à faixa média. Observamos que essas três palavras são monossílabas, então pode estar havendo um efeito do número de sílabas na suscetibilidade a hesitações. Conforme já afirmamos anteriormente, o grupo em *–l* é maior que a classe em ditongo em *–u*, no entanto há mais monossílabos em *–u* que em *–l*, talvez por isso os monossílabos favoreçam hesitações. O

efeito da variável número de sílabas será discutido no "Anexo B" (Tabela 62), uma vez que já comentamos esse fator quando analisamos os dados gerais sobre plurais em *–l*. Com relação aos itens que desfavorecem hesitações, observamos que "anzol" e "agradável" são de freqüência baixa, "anel" é de média e "responsável" é de alta freqüência de ocorrência. Portanto, para o grupo de plurais em *–l*, não houve, a princípio, correlação entre freqüência de ocorrência e favorecimento de hesitações.

Para prosseguir com a análise, removemos, também, o fator palavra, pelos mesmos motivos já mencionados anteriormente. Fazendo novas rodadas no SPSS, o programa considerou significativas as seguintes variáveis: **número de sílabas**, **vogal precedente, tonicidade** e **estrutura morfológica**. Na subseção 6.2, quando analisamos os resultados gerais para os plurais em *–l*, as variáveis número de sílabas, vogal precedente e estrutura morfológica já haviam sido exploradas (Cf. Tabelas 41, 39, 40, 37 e 38). Comentaremos, nesta análise de dados, apenas a tonicidade, porque essa variável não foi relevante nos resultados gerais para os plurais em *–l*. As tabelas e comentários para as demais variáveis significativas no "Teste de Reação" podem ser encontrados no "Anexo B" (Tabelas 62, 63 e 64). Vejamos, então, a tabela abaixo sobre tonicidade:

Tabela 59: Efeito da tonicidade no "Teste de Reação" para plurais em *–l*

Tonicidade	**N**	**%**	**P.R.**[99]
Oxítonas	77/466	16,5	0.59
Paroxítonas	21/180	11,6	0.40

Observamos que, nos plurais em *–l*, as palavras oxítonas favorecem as hesitações (0.59), ao passo que as paroxítonas as inibem (0.40). Esse é um efeito que pode, também, decorrer das similaridades fonológicas entre as classes de palavras em *–l* e ditongo em *–u*. A freqüência de tipo do padrão em *–l* é mais alta que a do paradigma em ditongo em *–u*, no entanto, para haver analogia entre diferentes classes, além da freqüência de tipo, é necessário que ocorra, também, atuação de similaridades fonológicas (Cf. BYBEE; SLOBIN, 1982; BYBEE; MODER, 1983; HARE; ELMAN, 1995). Uma palavra não pode migrar para outra classe se não houver, entre os dois grupos, traços de semelhanças em termos fonológicos. No caso da tonicidade, a classe de palavras terminadas em ditongo em

[99] No cálculo sobre peso relativo para tonicidade, removemos a variável vogal precedente, porque estava havendo interação entre esses dois fatores, fazendo com que os pesos relativos fossem muito distantes com relação aos valores percentuais.

–u é composta somente por itens oxítonos. Não há, nesse grupo, sequer uma palavra que seja paroxítona. Na classe de *–l*, por outro lado, há palavras paroxítonas e oxítonas. Esse talvez seja o grande responsável pelas hesitações nos plurais dos itens oxítonos. Já que, no singular, itens terminados em *–l* e *–u* são pronunciados como [w], não há distinção fonética que bloqueie a utilização de um plural por outro, portanto generalizações de uma classe podem afetar a outra. Quando encontram uma palavra oxítona, os falantes tentam alocá-la em seu espaço acústico a partir de similaridades com itens já estocados. Se na classe de ditongo em *–u* há apenas oxítonos, essa é uma generalização forte, capaz de atrair outros membros terminados em [w] que sejam, também, oxítonos. Essa é uma explicação possível para o fato de palavras oxítonas terminadas em *–l* favorecerem hesitações.

No grupo de plurais em *–l*, observamos que freqüência de ocorrência do item lexical não foi significativa para as hesitações no "Teste de Reação". Na análise global dos dados (Cf. subseção 6.2), para que os efeitos de freqüência de ocorrência fossem relevantes, foi necessário excluir as variáveis número de sílabas, vogal precedente e estrutura morfológica, porque, conforme vimos, essas foram as que mais interferiram na migração de plurais de *–l* para ditongo em *–u*. No caso do "Teste de Reação", excluímos, também, esses mesmos fatores, a fim de verificar se efeitos de freqüência de ocorrência poderiam ser observados. Contrariamente aos resultados gerais para plurais em *–l*, a exclusão dessas variáveis no "Teste de Reação" não fez com que a freqüência de ocorrência fosse considerada relevante para as hesitações. Tentamos fazer alguns outros ajustes, como remover as variáveis extralingüísticas (faixa etária, escolaridade e gênero), ou excluir a palavra "gol" (conforme fizemos na análise geral dos dados), mas em nenhuma dessas tentativas freqüência de ocorrência foi selecionada. Concluímos, então, que a variável freqüência de ocorrência não é relevante para que hesitações aconteçam no "Teste de Reação" para plurais em *–l*.

Como nossa tese defende que o léxico mental é organizado e remodelado a partir de efeitos de freqüência, é importante discutir por que, no "Teste de Reação" para palavras terminadas em *–l*, não houve nenhuma interferência da freqüência de ocorrência. Nossa hipótese para isso é a seguinte: conforme o Modelo de Redes postula, é a freqüência de tipo e não a de ocorrência que torna uma classe paradigmática mais forte no léxico mental. Sendo assim, em termos de tipos, o plural em *–l* é mais recorrente que o plural em ditongo em *–u* (Cf. Tabelas 8, 9, 10 e 11, no Capítulo 3, "Revisão de Literatura"). Por causa disso, as migrações de *–l* para *–u* são menores do que as de ditongo em *–u* para *–l*. Já que a

classe de palavras terminadas em *–l* é maior que a de palavras terminadas em ditongo em *–u*, mesmo que a palavra em *–l* seja pouco freqüente, sua classe é suficientemente forte e estável para prevenir seus itens (mesmo os de baixa freqüência de ocorrência) contra mudanças analógicas. A classe concorrente (de ditongo em *–u*) tem um número menor de itens e, por isso, não seria capaz de atrair membros do grupo em *–l*, já que esses têm maior freqüência de tipo[100]. Portanto, na classe de palavras em *–l*, a freqüência de tipo é a mais relevante, garantindo estabilidade aos itens léxicos, mesmo àqueles que, individualmente, apresentem baixa freqüência de ocorrência. Possivelmente, por causa disso, não houve efeitos de freqüência de ocorrência no "Teste de Reação" para palavras terminadas em *–l*.

Com relação aos plurais em ditongo em *–u*, nosso "Teste de Reação" contou com 607 dados, nos quais houve 115 hesitações, o que equivale a 18,9% dos dados. O grupo em ditongo em *–u*, foi, portanto, o mais suscetível a hesitações. Uma possível explicação para isso pode ser o fato de que a classe em ditongo em *–u* é, dentre os três grupos analisados nesta tese, a que apresenta menor freqüência de tipo, tanto no ASPA quanto no Novo Dicionário Aurélio (Cf. Tabelas 5, 6, 8, 9, 10 e 11, no Capítulo 5, "Metodologia"). Os grupos em *–ão* e *–l* são mais freqüentes que o de ditongo em *–u*, talvez por isso tenham sido menos suscetíveis a hesitações. Quando vão pronunciar os plurais de palavras terminadas em *–u*, os informantes apresentam dúvidas provavelmente porque a baixa freqüência de tipo desses itens faz com que não sejam recuperados prontamente no léxico mental. Sendo assim, para prover esses plurais, o falante hesita, enquanto tenta acessar a forma adequada em seu léxico.

Submetendo os dados sobre o "Teste de Reação" de plurais em ditongo em *–u* ao SPSS, as variáveis **palavra** e **escolaridade** foram selecionadas. Vejamos, abaixo, os resultados para as **palavras** individualmente:

---

[100] Exceção deve ser feita quanto os itens monossílabos e oxítonos terminados em ditongo em *–u*. Como vimos anteriormente, a classe em ditongo em *–u* é, como um todo, menor que a de *–l*; no entanto, apresenta maior número de monossílabos e oxítonos e, por conta dessa similaridade fonológica, é capaz de atrair membros também monossílabos e oxítonos que sejam da classe em *–l*.

Tabela 60: Efeito da palavra no "Teste de Reação" para plurais em ditongo em *–u*

Palavras	Freqüência	N	%	P.R.
1. Jirau	**baixa – 01**	**18/35**	**51,4**	**0.85**
2. Mausoléu	**baixa – 04**	**19/35**	**54,2**	**0.87**
3. Ateu	baixa – 24	4/35	11,4	0.41
4. Véu	**baixa – 33**	**12/36**	**33,3**	**0.73**
5. Pau	baixa – 46	7/36	19,4	0.57
6. Troféu	baixa – 72	3/35	8,5	0.34
7. Degrau	média – 109	4/36	11,1	0.41
8. Chapéu	média – 125	6/36	16,6	0.52
9. Céu	média – 163	3/36	8,3	0.33
10. Réu	média – 309	7/36	19,4	0.57
11. Mau	média – 346	5/36	13,8	0.47
12. Museu	média – 424	4/36	11,1	0.41
13. Judeu	**alta – 808**	**12/35**	**34,2**	**0.74**
14. Grau	alta – 916	5/36	13,8	0.47
15. Europeu	alta – 1.119	4/36	11,1	0.41
16. Meu	*alta – 2.139*	*1/36*	*2,7*	*0.13*
17. Seu	*alta – 29.682*	*1/36*	*2,7*	*0.13*

Na tabela acima, observamos, primeiramente, que todas as palavras apresentaram pelo menos um caso de hesitação. Nos dados sobre os plurais em *–ão* e *–l*, percebemos que alguns itens não causaram nenhuma dúvida nos falantes. No caso de plurais em ditongo em *–u*, houve dúvidas em todas as palavras adotadas no "Teste de Reação", talvez justamente por esse ser, entre os grupos analisados nesta tese, aquele que apresenta a menor quantidade de tipos. Em negrito destacamos os itens mais suscetíveis a hesitações e, em itálico, os menos. As palavras que favoreceram hesitações foram "mausoléu", "jirau", "judeu" e "véu". Desse grupo, "judeu" é de freqüência de ocorrência alta e as outras três, de baixa ocorrência. De acordo com o Modelo de Redes e a Teoria dos Exemplares, o fato de que itens de baixa freqüência de ocorrência causem hesitações era previsto, mas não era de se esperar que o item "judeu" (de alta freqüência de ocorrência) também provocasse dúvidas nos falantes. Uma hipótese para que essa palavra tenha sido uma das líderes nas hesitações é o fato de que, apesar de freqüente, ela não é familiar para os informantes consultados em nossos experimentos. Esse item é freqüente no Corpus NILC/São Carlos, que contém dados de escrita; no entanto, na linguagem cotidiana, pode ser que os falantes não usem a palavra "judeu" com a mesma recorrência com que ela

aparece em um corpus de escrita. Por causa disso, quando são questionados sobre o plural desse item, os falantes tendem a hesitar. A discrepância entre freqüência de ocorrência em um corpus de língua e familiaridade para o informante individualmente pode, portanto, justificar por que "judeu", apesar de sua alta freqüência de ocorrência, é uma das palavras que mais causaram hesitação no "Teste de Reação".

As duas palavras que inibiram dúvidas foram "meu" e "seu", justamente as duas mais freqüentes de nossos dados. Observamos, então, que, no grupo de palavras terminadas em ditongo em *–u*, existe correlação entre freqüência de ocorrência e propensão a hesitações nos plurais. Quanto maior a freqüência de ocorrência, mais disponível o item estará no léxico mental e, conseqüentemente, menores serão as dúvidas para formar seu plural; contrariamente, quanto menos freqüente a palavra, menos acessível será no léxico mental e mais dependerá de sua classe para que seja flexionada. Sendo a classe (em ditongo em *–u*) também de baixa freqüência de tipo, as hesitações são maiores nesse grupo, principalmente nas palavras de baixa freqüência de ocorrência.

Com relação à variável escolaridade, também selecionada nesta primeira rodada do SPSS, faremos comentários sobre ela no "Anexo B" (Tabela 67), já que, na análise total dos dados, já havíamos mencionado os efeitos desse fator no grupo de plurais em ditongo em *–u*.

Prosseguindo com a análise do "Teste de Reação" para plurais em ditongo em *–u*, excluímos a variável palavra, visto que já havíamos feito os comentários pertinentes a ela e precisávamos verificar se, removendo esse fator, haveria seleção de outras variáveis significativas para hesitações. Sem a variável palavra, o SPSS considerou relevantes os seguintes fatores: **freqüência de ocorrência**, **estrutura morfológica**, **vogal precedente** e **escolaridade**. Esses fatores já haviam sido selecionados quando analisamos os dados gerais para os plurais de ditongo em *–u* (Cf. Tabelas 51, 49, 48 e 53). Para evitar redundâncias, não vamos retomar, aqui, as mesmas variáveis já comentadas na subseção sobre os plurais em ditongo em *–u*. Mencionaremos apenas os resultados para freqüência de ocorrência, já que esse é o foco principal de nossa pesquisa. Os resultados para as demais variáveis consideradas significativas pelo SPSS são apresentados no "Anexo B" (Tabelas 65, 66 e 67). Vejamos, portanto, a tabela abaixo, que analisa o papel da freqüência de ocorrência no "Teste de Reação" sobre plurais em ditongo em *–u*:

Tabela 61: Efeito da freqüência de ocorrência no "Teste de Reação" para plurais em ditongo em *–u*

Faixas de Freqüência	N	%	P.R.
Freqüência Baixa	63/212	29,7	0.69
Freqüência Média	29/216	13,4	0.44
Freqüência Alta	23/179	12,8	0.35

No "Teste de Reação" para as palavras terminadas em ditongo em *–u*, houve uma gradualidade dos pesos relativos com relação à freqüência de ocorrência. Observamos, então, que as palavras pouco freqüentes favoreceram as hesitações (0.69), as de média freqüência desfavoreceram um pouco as hesitações (0.44) e as de alta freqüência desfavoreceram ainda mais as hesitações (0.35). Então, as dúvidas dos falantes foram proporcionais à freqüência de ocorrência das palavras.

No caso dos plurais em ditongo em *–u*, todas essas palavras apresentam um paradigma de pluralização regular na língua, já que seu plural é formado através da adição do morfema *–s*, o mais comum do PB. Sendo assim, a Teoria de Palavras e Regras não preveria que efeitos de freqüência de ocorrência fossem encontrados nessa classe (Cf. Capítulo 4, "Quadro Teórico"). No entanto, nossos dados demonstram que, mesmo entre os itens regulares (terminados em ditongo em *–u*), houve efeitos de freqüência de ocorrência, da seguinte forma: palavras mais freqüentes tiveram acesso mais rápido no léxico mental, ao passo que palavras pouco freqüentes foram acessadas mais lentamente. Os resultados de nossos experimentos desafiam, portanto, os postulados da Teoria de Palavras e Regras sobre o fato de que apenas itens irregulares são suscetíveis a efeitos de freqüência de ocorrência.

Comentando os resultados do "Teste de Reação" como um todo, observamos que nossas hipóteses de trabalho foram comprovadas, porque os plurais em *–ão* e em ditongo em *–u* (grupos em que a migração de plurais efetivamente se dá) foram suscetíveis a efeitos de freqüência de ocorrência. Nessas classes, falantes demonstraram mais hesitação para pronunciar o plural de palavras infreqüentes e proveram respostas mais rápidas quando as palavras eram freqüentes, o que demonstra que a freqüência de ocorrência é um grande determinante para o acesso a itens no léxico mental.

O Modelo de Redes e o de Exemplares postulam que a autonomia lexical é conseqüência da alta freqüência de ocorrência. Se um item é altamente freqüente, está listado no léxico e seu acesso é feito de forma direta e mais rápida, sem que seja necessário

recorrer à sua classe; por outro lado, se um item tem baixa freqüência de ocorrência, é acessado através de sua classe, o que faz com que demore mais para ser flexionado. Devido a esses motivos, os falantes proveram com mais rapidez o plural das palavras freqüentes e mais lentamente o plural das infreqüentes. No grupo de palavras terminadas em *–l*, tais efeitos não são observados porque a freqüência de tipo dessa classe é maior que a da classe concorrente (em ditongo em *–u*), então isso, de certa forma, impede que haja migrações de plurais em *–l* para ditongo em *–u*, reduzindo, conseqüentemente, a taxa de hesitações nas respostas. Mesmo que o acesso a itens infreqüentes terminados em *–l* seja feito através da rede, não da palavra individualmente, esse já é o tipo mais freqüente para itens terminados em [w], por isso o acesso via classe não aumenta o tempo de ativação da palavra.

Na subseção seguinte, faremos uma discussão mais ampla sobre os resultados alcançados em nossos experimentos.

6.5 Discussão geral dos resultados

Iniciando esta subseção em que propomos uma análise mais geral acerca dos resultados para todos os grupos de plurais, vamos retomar a discussão sobre as teorias de léxico mental que esta tese se propõe a investigar, contrapondo modelos que pressupõem regras abstratas com modelos que negam a necessidade de regras e defendem a emergência de generalizações a partir de itens léxicos (regulares ou irregulares) estocados.

Em linhas gerais, a Teoria de Palavras e Regras postula que há dois caminhos de acesso a palavras no léxico mental: o primeiro é através de memorização para itens irregulares e, portanto, não deriváveis a partir de regras; o segundo é através da aplicação de regras para itens regulares. Sendo assim, quando vai pronunciar uma palavra, o falante, em primeiro lugar, tentaria buscar esse item em seu léxico mental; não o encontrando, é sinal de que ele não está lexicalmente estocado porque é regular e, portanto, a regra canônica, mais recorrente da língua ("default") se aplica. Nessa teoria, efeitos de freqüência de ocorrência são previstos apenas para itens irregulares, já que apenas eles dependem de memória para que sejam relembrados pelo falante. Palavras regulares podem ser derivadas

por regras, não requerem memorização e, conseqüentemente, não seriam suscetíveis a efeitos de freqüência.

De forma contrária a essa posição, o Modelo de Redes e a Teoria dos Exemplares assumem que palavras inteiras podem ser listadas no léxico mental e que a freqüência (de tipo e/ou de ocorrência) tem papel crucial na representação mental de informações lingüísticas. Sendo assim, as palavras estão estocadas no léxico mental em redes de interconexões fonológicas e semânticas. Propriedades morfológicas, em vez de definidas *a priori*, emergem de generalizações a partir dessas redes de similaridades. Nesse sentido, o léxico mental está sempre sendo atualizado no dia-a-dia dos falantes, à medida que novos exemplares são aprendidos e outros, mais antigos, vão caindo no esquecimento.

Analisando os plurais do PB à luz dessa discussão teórica, observamos que os dados apresentam contribuições importantes para o debate sobre como o léxico mental categoriza, armazena e acessa informações lingüísticas. Retomando as hipóteses dos modelos teóricos para os grupos de plurais, a Teoria de Palavras e Regras esperaria que, com relação ao grupo de plurais em *–ão*:

- ✓ As palavras com plural etimológico em *–ãos* preservassem suas formas, já que a pluralização regular, através da adição do morfema *–s* (*–ão* + *–s*) , é a mais recorrente na língua;
- ✓ As palavras com desinência *–ões* e *–ães* etimológicas fossem estocadas no léxico, por conta de sua irregularidade, e os itens em *–ãos* (*–ão* + *–s*) não fossem listados, já que são regulares e deriváveis por regras abstratas, prescindindo da memorização. Sendo assim, por conta da estrutura regular ou irregular dos itens léxicos, somente para os grupos em *–ões* e *–ães* etimológicos deveria haver efeitos de freqüência de ocorrência. Os plurais em *–ãos*, visto que são derivados por regras e não estão listados no léxico mental, não deveriam ser influenciados pela freqüência de ocorrência.

Quanto aos plurais em *–l* e ditongo em *–u*, a Teoria de Palavras e Regras suporia que:

- ✓ Os plurais em ditongo em *–u*, já que são formados a partir da regra canônica da língua (adição de *–s*), deveriam prevalecer;

- ✓ Efeitos de freqüência de ocorrência poderiam ser encontrados apenas na classe em *–l*, já que essa é irregular e depende da memorização de seus itens léxicos, podendo, então, ser afetada pela freqüência da palavra individualmente. No grupo de palavras terminadas em ditongo em *–u*, visto que a pluralização é regular, as palavras não estariam estocadas no léxico mental e, portanto, efeitos de freqüência de ocorrência não deveriam acontecer.

Por sua vez, o Modelo de Redes e o de Exemplares formulariam hipóteses diferentes para os grupos de plurais analisados nesta tese. Com relação aos plurais em *–ão*, tais teorias suporiam que:

- A freqüência de tipo do padrão em *–ões* suplantaria a pluralização regular da língua (em *–s*) e atrairia itens pertencentes às classes etimológicas em *–ãos* e *–ães* (tipos menos freqüentes);
- Palavras de alta freqüência de ocorrência, ainda que pertençam aos tipos *–ãos* e *–ães* (pouco freqüentes), deveriam preservar seus plurais etimológicos, já que são mais autônomas no léxico por conta de sua freqüência individualmente; palavras com tipos infreqüentes (*–ãos* e *–ães*) e também com baixa freqüência de ocorrência deveriam favorecer a adoção de plurais em *–ões*, já que não têm exemplares suficientemente fortes no léxico mental e, por isso, tendem a adotar a classe paradigmática mais representativa em termos de tipos (no caso, *–ões*).

Para os plurais em *–l* e ditongo em *–u*, as hipóteses do Modelo de Redes e da Teoria dos Exemplares seriam as seguintes:

- Como não há distinção fonética entre *–l* e *–u* em fim de sílaba no PB, os plurais em *–is* e *–us* são ambos possíveis para palavras que terminam em [w] (*–l* ou *–u* ortográficos). Sendo *–l* em fim de palavra um tipo mais freqüente que ditongo em *–u* (na mesma posição), as palavras terminadas em *–u* poderiam migrar para a classe em *–l*, sendo atraídas por sua alta freqüência de tipo;
- Já que a freqüência de ocorrência interfere tanto em itens regulares quanto irregulares, as palavras em ditongo em *–u* com baixa freqüência de ocorrência seriam as primeiras a migrar para o padrão em *–l*; por outro lado, palavras em *–u*

com alta freqüência de ocorrência tornar-se-iam mais autônomas no léxico mental e poderiam preservar plurais de tipos infreqüentes (no caso, ditongo em *–u*).

Conforme vimos ao longo deste capítulo, os resultados alcançados com nossa análise são mais compatíveis com o Modelo de Redes e o de Exemplares do que com a Teoria de Palavras e Regras. A preferência dos falantes pelo morfema *–ões* (em relação a *–ãos* e *–ães*) foi recorrente, e, no grupo de ditongo em *–u* (em que a pluralização em *–s* deveria prevalecer), houve maior probabilidade de migração para o grupo em *–l* do que de palavras em *–l* migrando para a classe de ditongo em *–u*.

Com relação aos processos de pluralização, na classe em *–ão*, os falantes adotaram mudança de radical, transformando *–ão* em *–ões*, em vez de apenas utilizarem a flexão através do morfema *–s*, mesmo sendo este (em *–s*) o mais freqüente de todos os padrões de pluralização de nossa língua. No grupo de ditongo em *–u*, os falantes fizeram o mesmo, transformando *–au* e *–eu* (ou *–éus*) em *–ais* e *–eis* (ou *–éis*), ou seja, mudando o radical da palavra, em vez de simplesmente adotarem a regra mais produtiva da língua, a flexão com o morfema *–s*. Como Bybee e Newman (1995) observaram, sufixação não é mais fácil do que mudança de radical. O que realmente conta é a freqüência de cada um desses processos no uso da língua. Sendo os tipos em *–ões* e *–is* mais freqüentes que *–ãos*/ *–ães* e *–us*, respectivamente, os informantes adotaram tais pluralizações (em *–ões* e *–is*), em vez de optarem pela regra geral da língua. Isso parece sugerir que, quando os falantes têm de pluralizar uma palavra, não consideram somente o segmento final (vogal ou consoante) ou a regra geral de pluralização em *–s*. Em vez disso, eles avaliam a palavra inteira e a freqüência de tipo para cada uma das possibilidades de plural. A freqüência de tipo do padrão regular em *–s* é a maior da língua; apesar disso, os falantes demonstraram, em parte dos dados (Cf. Tabelas 25 e 34), preferência pela mudança de radical, adotando *–ões* para *–ão* e *–is* para ditongo em *–u*. Nesse caso, Bybee (1995) sugere que regularidade e produtividade estão dissociadas, uma vez que o padrão mais regular de plural na língua (em *–s*) não é o mais produtivo para todas as classes de palavras terminadas em vogal.

O fato de os falantes, ao pluralizarem uma palavra, considerarem não somente o segmento final, mas a palavra como um todo e a freqüência das desinências de plural, é um forte argumento contra a existência de regras no léxico mental. Se os falantes usassem apenas regras na pluralização das palavras, eles adotariam *–s* (o morfema mais freqüente de plural) em todas as pluralizações regulares, transformando *–ão* em *–ãos* e *–u* em *–us*. Em vez

de usarem regras, os falantes parecem basear-se em esquemas locais formados a partir de similaridades entre os itens. Talvez essa seja uma evidência de que as palavras estão estocadas no léxico como unidades inteiras e de que generalizações morfológicas sobre classes emergem a partir das conexões entre essas palavras. Pluralizar uma palavra é, portanto, não apenas aplicar uma regra para um grupo de palavras, mas tomar uma decisão lexical entre itens listados no "dicionário" mental.

Os dados analisados sugerem que os falantes pluralizam as palavras a partir de informações probabilísticas sobre desinências de plurais e generalizações locais sobre grupos de itens léxicos. Segundo o Modelo de Redes e a Teoria dos Exemplares, palavras inteiras estão listadas no léxico e, por isso, o acesso à base (item primitivo) e ao produto (forma derivada) se dá simultaneamente. Sendo assim, o acesso a uma forma plural reforça a singular e vice-versa. O fato de os falantes não demonstrarem preferência pela pluralização em *–s* (tipo mais freqüente de todos os plurais do PB) para a pluralização de *–ão* e ditongo em *–u* demonstra que há esquemas locais determinados por grupos de palavras e, nesses casos, existem padrões específicos mais freqüentes que se aplicam por conta de similaridades fonológicas.

No caso do grupo em *–ão*, o fato de todas as palavras no singular terminarem da mesma forma (em *–ão*) parece ser uma forte referência para que os falantes saibam que há, potencialmente, três possibilidades de plural (*–ões*, *–ãos* e *–ães*). Ao encontrarem um novo item em *–ão*, os falantes tentam alocá-lo no espaço acústico de seu léxico mental, de acordo com as semelhanças entre esse novo item e os antigos, já estocados. A similaridade fonológica entre os membros dessa classe (talvez baseada na presença da vogal nasal /ã/) faz com que ela não seja confundida com um grupo de itens regulares que termina em vogal e seria pluralizado através do morfema *–s*. A partir das redes em seu léxico mental, os falantes parecem abstrair o fato de que, dentre os três morfemas concorrentes (*–ões*, *–ãos* e *–ães*), *–ões* é o mais freqüente e, portanto, deve ser utilizado quando o plural da palavra não é relembrado por causa de sua baixa freqüência de ocorrência. Na pluralização geral da língua, o padrão em *–s* é o mais recorrente, mas *–ões* é específico para pluralizar itens terminados em *–ão*, é o mais freqüente nessa classe e vem atraindo itens diacronicamente. Além disso, se o acesso a itens primitivos e derivados ocorre simultaneamente, a conexão lexical entre *–ão* e *–ões* se fortalece, tornando-se mais proeminente no léxico mental e fazendo com que *–ões* tenda a ser aplicado todas as vezes que o falante precisa pluralizar um item cuja forma, devido à sua baixa freqüência de ocorrência, não esteja listada no léxico. É

claro que a pluralização em *–ões* é limitada se for comparada ao plural em *–s*, no entanto a especificidade de *–ões* (usado apenas para palavras terminadas em *–ão*) garante sua produtividade. Nesse sentido, a alta freqüência de tipo do morfema *–ões* para o padrão em *–ão* no singular está se sobrepondo ao padrão mais freqüente de plural do PB, que é a pluralização em *–s*.

Em termos etimológicos, a pluralização de *–ano*, *–ane* e *–one* do latim era regular, feita através do acréscimo do morfema *–s*. Com o nivelamento das formas singulares em direção a *–ão*, as palavras estariam prontas para adotar o plural regular da língua (em *–s*). No entanto, *–ões*, desde o português arcaico, já pluralizava uma quantidade maior dos itens que se fundiram em *–ão* (Cf. Capítulo 2, "Grupos de Plurais"). Por causa disso, apresentava, conseqüentemente, mais palavras listadas no léxico e essa alta freqüência de tipo fez com que se tornasse o padrão de plural para a classe terminada em *–ão* no singular. Portanto, a despeito da regra canônica de pluralização do PB ser o morfema *–s*, no grupo em *–ão*, o plural produtivo é em *–ões*. O Modelo de Redes e a Teoria dos Exemplares postulam que palavras de alta freqüência de ocorrência são acessadas como unidades inteiras, por isso não dependem de sua classe para que sejam flexionadas. Por outro lado, se a palavra é infreqüente, sua baixa freqüência de ocorrência não garante sua estocagem lexical e ela pode ser acessada através de sua classe. Sendo *–ões* o morfema mais freqüente para *–ão* no singular, quando os falantes se deparam com uma palavra infreqüente terminada em *–ão*, acabam por alocá-la na classe de itens em *–ões*, por ser esta terminação a mais freqüente no plural. Como *–ãos* e *–ães* são infreqüentes (em termos de tipo), não têm força de classe lexical, por isso só sobrevivem nos itens que, individualmente, apresentam alta freqüência de ocorrência e se tornam autônomos no léxico mental, prescindindo, portanto, da referência de sua classe para serem relembrados. Tais resultados foram comprovados através da análise de dados como um todo, já que itens em *–ãos* e *–ães* etimológicos de baixa freqüência de ocorrência favoreceram plurais em *–ões*, e também no "Teste de Reação", visto que os falantes hesitaram menos quando o plural etimológico era *–ões* (tipo mais freqüente) e hesitaram mais em palavras de baixa freqüência do que de alta freqüência (pertencentes a *–ões*, *–ãos* ou *–ães*), demonstrando efeitos de freqüência no acesso a itens regulares e irregulares.

No caso das palavras terminadas em *–l* e ditongo em *–u*, efeitos de freqüência também foram encontrados. Segundo o Modelo de Redes e a Teoria dos Exemplares, não há distinção categórica entre regular e irregular, produtivo ou improdutivo, porque essa

divisão reflete, na verdade, dois extremos de um mesmo *continuum*. Sendo assim, o fato de os itens terminados em ditongo em *–u* serem regulares (pluralizados através do morfema *–s*) não os previne contra generalizações decorrentes do uso. Nossos dados demonstraram que houve, também, efeitos de freqüência no grupo em ditongo em *–u*, com palavras de baixa freqüência de ocorrência favorecendo o plural em *–l*. Observamos que, perdida a distinção entre *–l* e *–u* em fim de sílaba, não há barreira para que a pluralização em *–l* ou ditongo em *–u* se aplique aos itens terminados em [w] no singular. Como vimos no Capítulo 3, "Revisão de Literatura", o tipo *–l* em fim de sílaba é mais freqüente que o tipo em ditongo em *–u* (também nessa posição). Já que não há uma distinção fonética que separe categoricamente as duas classes (em *–l* e *–u*), a alta freqüência do tipo em *–l* está atraindo membros do grupo em *–u*. A similaridade fonética entre as duas classes permite que esse tipo de generalização ocorra e se difunda na língua. Segundo Hare e Elman (1995), para que aconteça mudança de classe analógica, é necessário que haja, primeiramente, identidade fonética entre os itens pertencentes às duas classes. Esse parece ser o caso exato dos itens em *–l* e ditongo em *–u*. Foneticamente esses dois grupos formam uma classe só (com [w] final) e, portanto, a pluralização pode ser em *–is* ou em *–us*. Sendo *–is* mais freqüente que *–us*, palavras terminadas em ditongo estão adotando plural em *–l*.

Paralelamente a esse efeito de freqüência de tipo, observamos que a freqüência de ocorrência também afeta essas palavras (em *–l* e em *–u*). Assim, palavras da classe em ditongo em *–u*, quando apresentam alta freqüência de ocorrência, tendem a ser preservadas, porque sua ocorrência individual faz com que sejam listadas no léxico e não precisem ser acessadas através de seu grupo; por outro lado, palavras terminadas em *–u* que apresentam baixa freqüência de ocorrência são as primeiras a migrar para a classe em *–l*, porque sua baixa ocorrência não lhes permite armazenamento individual no léxico. Para serem lembradas, dependem do acesso à sua rede e, como esta apresenta baixa freqüência de tipo, não garante que o plural seja preservado. Assim, palavras em ditongo em *–u* de baixa freqüência de ocorrência lideram a adoção de plural analógico em *–l*. Mais uma vez, esses resultados foram verificados na análise geral dos dados e no "Teste de Reação". Na análise total, observamos que a freqüência de ocorrência foi determinante para que as palavras (tanto em *–l* quanto em ditongo em *–u*) preservassem seus plurais ou migrassem para a classe concorrente. No "Teste de Reação", percebemos que a baixa freqüência de ocorrência gerou hesitações no grupo de plurais em ditongo em *–u*, mas, na classe em *–l*, a freqüência de ocorrência não foi significativa, justamente porque o tipo em *–l* já é o mais

freqüente (comparado a ditongo em *–u*) e, portanto, o acesso às palavras desse grupo é feito de forma mais rápida.

Uma comparação cabível entre a classe em *–ão* e os grupos em *–l* e ditongo em *–u* é a seguinte: diacronicamente, o processo de prevalência de *–ões* sobre *–ãos* e *–ães* se iniciou a partir da fusão das formas singulares latinas (Cf. Capítulo 2, "Grupos de Plurais"). Não havendo distinção fonética no singular (todas as formas se tornaram *–ão*), qualquer uma das três terminações (*–ões*, *–ãos* ou *–ães*) poderia ser adotada. O morfema *–ões* prevaleceu porque era aplicado a um número maior de itens, ou seja, possuía freqüência de tipo mais alta já no português arcaico. No caso de *–l* e ditongo em *–u*, estamos observando, sincronicamente, um fenômeno semelhante: não há mais distinção fonética no singular (as duas classes de palavras terminam em [w]) e, por isso, as formas plurais tanto em *–is* quanto em *–us* são válidas para os itens desses dois grupos. Já que o tipo *–is* é o mais freqüente no plural, pode ser que este, futuramente, seja adotado de forma mais ampla para todos os itens singulares em *–l* e ditongo em *–u*. A ortografia é, atualmente, o único correlato que pode garantir a preservação das formas plurais etimológicas para os grupos em *–l* e ditongo em *–u*, no entanto é de se questionar se uma distinção ortográfica poderá impedir uma fusão de classes que perderam a distinção fonética. Em um tempo futuro, pode ser que ocorra, nos grupos em *–l* e ditongo em *–u*, o mesmo fenômeno que aconteceu com os plurais em *–ão*.

Retomando a análise de dados de uma forma mais ampla, observamos que, nos três grupos de plurais analisados, houve casos de generalizações locais capazes de prevenir a adoção de plurais analógicos. Comentaremos esses casos em separado.

No grupo de palavras terminadas em *–ão*, observamos que itens monossílabos e palavras paroxítonas desfavoreceram a adoção do plural analógico em *–ões* e tenderam a preservar seus plurais etimológicos, em *–ãos* ou *–ães*, apesar de sua baixa freqüência de tipo. Esse resultado demonstra que, conforme Bybee e Slobin (1982) afirmam, existem esquemas locais formados por similaridades fonológicas e semânticas entre as palavras. Nesses esquemas, a repetição de um padrão específico (de plural para uma determinada classe) se convencionaliza e se torna obrigatória em certas condições (monossílabos se pluralizam em *–ãos* e *–ães*, paroxítonos se pluralizam em *–ãos*). Já que os itens estão listados no léxico como unidades inteiras, esses esquemas são deriváveis das interconexões que existem entre as palavras e, por isso, plurais de tipos infreqüentes podem ser mantidos em um local específico, como nos exemplos que fornecemos para o grupo em *–ão*. Observa-se,

portanto, a existência de conhecimento detalhado e probabilístico sobre cada um dos morfemas de plural e sobre o grupo de palavras a que eles, preferencialmente, se aplicam.

Nas classes de palavras terminadas em *–l* e ditongo em *–u* (consideradas em conjunto), observamos outras generalizações: palavras monossílabas em *–l* favoreceram plural em ditongo em *–u* e palavras polissílabas em ditongo em *–u* favoreceram plural em *–l*; palavras com sufixo favoreceram seus plurais etimológicos nas duas classes. A generalização quanto ao número de sílabas sugere que, conforme o Modelo de Exemplares afirma, as palavras são categorizadas em termos de exemplares prototípicos e periféricos. A partir de informações sobre os padrões probabilísticos disponíveis no léxico, o falante constrói o conhecimento de que, no grupo em *–l*, o protótipo de palavra é polissílaba, ao passo que, no grupo em ditongo em *–u*, o protótipo é palavra monossílaba. Itens periféricos também compõem a nuvem de exemplares, porque partilham alguma semelhança (como a terminação em [w]) com os membros centrais, por isso há, na classe em *–l*, itens monossílabos, e, no grupo em ditongo em *–u*, itens com mais de uma sílaba, embora esses sejam em menor número. Já que não há distinção fonética entre *–l* e *–u* em fim de sílaba, os falantes tendem a fundir essas duas classes em uma só, aplicando a elas padrões de generalização decorrentes da comparação entre os dois grupos: monossílabos de ambos os grupos foram pluralizados como ditongo em *–u* e polissílabos (também dos dois grupos), como *–l*.

Com relação à estrutura morfológica, a existência de sufixo foi, nas duas classes (em *–l* e em *–u*), uma forte referência para a manutenção dos plurais etimológicos. Mais uma vez, observamos que a presença de sufixo em vários itens léxicos torna essa distinção marcante para seu grupo, evitando possíveis efeitos analógicos decorrentes de freqüência de tipo.

Em decorrência desses esquemas de generalizações locais, padrões infreqüentes (como *–ãos*, *–ães* e *–us*) podem ser mantidos, evitando a erosão de suas classes como um todo. A partir das conexões fonológicas e semânticas entre os itens armazenados no léxico do falante, redes para cada um dos grupos de plurais são formadas e informações morfológicas emergem. Dentro dessas redes, deve haver generalizações ainda mais específicas, que fortalecem a existência de esquemas particulares, aplicáveis a um subconjunto de itens dentro da classe de plural como um todo. A partir da proeminência desses esquemas particulares, alguns tipos lingüísticos de baixa freqüência podem ser mantidos em grupos específicos de palavras, como, por exemplo, monossílabos,

paroxítonos ou com sufixo. Quando os falantes são questionados sobre os plurais de palavras desses grupos específicos, as similaridades com outras palavras já estocadas em seu léxico mental (monossílabas, paroxítonas ou com sufixo, para retomar o exemplo acima) fazem com que esses novos itens sejam alocados nessa rede particular e recebam a pluralização peculiar à sua classe. Dessa maneira, os tipos pouco freqüentes mantêm suas irregularidades por causa da força dos esquemas locais.

Comentando, particularmente, os efeitos de freqüência de tipo, postulamos que as palavras inteiras estão listadas no léxico e a associação semântica e fonológica entre elas forma um esquema. Sendo *–ões* (para itens singulares em *–ão*) e *–is* (para itens singulares em [w]) muito produtivos (atraindo novos membros), acabam formando esquemas mais fortes. Os demais morfemas (*–ãos*, *–ães* e *–us*) são freqüentes em palavras individuais, mas a alta freqüência de ocorrência de itens específicos não contribui para a proeminência de sua rede. Uma classe se fortalece através de palavras de média e baixa freqüência, já que essas, por não terem alta freqüência de ocorrência, são relembradas mais facilmente através de seu paradigma. Portanto, sempre que são usadas, acessam, concomitantemente, sua rede, fortalecendo-a como um todo. Por outro lado, itens de alta freqüência de ocorrência (como "alemães", "mãos" e "museus") são autônomos no léxico mental e não dependem de seu grupo para que sejam rememorados. Seu acesso se dá de forma independente, por isso não fortalecem suas classes e não contribuem para que esses grupos sejam capazes de manter em seu paradigma os itens de baixa freqüência de ocorrência.

Além da freqüência de tipo, a freqüência de ocorrência foi, também, bastante significativa no que diz respeito às generalizações de plural, mesmo em palavras regulares (pluralizadas em *–ãos* ou *–us*). De acordo com a Teoria de Palavras e Regras, freqüência de ocorrência deveria afetar somente palavras irregulares, mas, em nossos dados, houve tais efeitos inclusive em itens regulares. Palavras com plural etimológico regular, como "vulcão" /"vulcão+s" e "mausoléu"/"mausoléu+s" foram pluralizadas como "vulc<u>ões</u>" e "mausolé<u>is</u>", demonstrando que também as palavras terminadas em vogal (regulares para a aplicação do padrão em *–s*) foram suscetíveis a efeitos de freqüência.

Se as palavras não estivessem listadas como unidades inteiras no léxico mental (como a Teoria de Palavras e Regras defende), como poderíamos explicar efeitos de freqüência de tipo e de ocorrência em itens regulares? Nossos dados mostram que é a freqüência de uso da palavra, não sua estrutura interna (regular ou irregular), que determina

sua estocagem no léxico mental. Sendo assim, a propensão à adoção de plurais analógicos ocorre em conseqüência do uso da palavra, não de sua (ir)regularidade.

De um modo geral, os resultados obtidos com nossos experimentos demonstram que a melhor explicação para favorecimento ou imunidade a generalizações de plural reside em uma interação entre freqüência de tipo e de ocorrência, nos seguintes termos: a alta freqüência de tipo torna uma classe paradigmática mais forte, por isso itens menos freqüentes tendem a ser atraídos para esse tipo de grupo, fazendo com que palavras de baixa freqüência de ocorrência se regularizem em direção a classes analógicas mais fortes. No entanto, se uma palavra tem alta freqüência de ocorrência e se flexiona através de um tipo de baixa freqüência, ela se torna independente de sua classe, ganha autonomia lexical e tal padrão tende a se manter nessa palavra, mesmo que sua classe paradigmática tenda a adotar o tipo mais recorrente na língua. Observamos, através de nossos dados, que tipos infreqüentes foram preservados em palavras de alta freqüência de ocorrência, como, por exemplo, "alemão"/"alemães", "europeu"/"europeus". Além disso, os resultados de nossos experimentos demonstram que os efeitos analógicos nas classes paradigmáticas são graduais, atingindo o léxico de forma sutil, começando, primeiramente, pelas palavras de baixa freqüência. Sendo assim, nos grupos terminados em *–ão* e ditongo em *–u*, são as palavras com pouca freqüência de ocorrência as primeiras a sofrerem efeitos analógicos em direção a padrões mais freqüentes (*–ões* e *–is*, respectivamente): "cristão" e "guardião" foram pluralizados como "cristões" e "guardiões", em vez de "cristãos" e "guardiães" (conforme a gramática padrão prevê); "jirau" e "mausoléu" foram pluralizados como "jirais" e "mausoléis", a despeito de "jiraus" e "mausoléus" serem as formas padronizadas. Assim, a baixa freqüência de ocorrência desses itens lexicais faz com que eles sejam os primeiros a sucumbir a mudanças analógicas. Os dados apresentam um comportamento compatível com os pressupostos da Difusão Lexical (WANG; CHENG, 1977; HSIEH, 1977; OLIVEIRA, 1997), no sentido de que a variação lingüística atinge o léxico gradualmente, afetando as palavras uma a uma. Não há, necessariamente, uma regularidade nesse processo, visto que nem todas as palavras adotam formas inovadoras na mesma velocidade. Em nossos dados, observamos essa gradualidade na maneira como os itens léxicos migram, individualmente, para classes analógicas de plurais.

Na subseção seguinte, finalizamos as conclusões sobre este capítulo.

6.6 Conclusões

Neste capítulo, apresentamos os resultados estatísticos para os dados coletados sobre os grupos de plurais em *–ão*, *–l* e ditongo em *–u*. Nossa interpretação dos resultados à luz do Modelo de Redes e da Teoria dos Exemplares concluiu que os efeitos de freqüência de tipo e de ocorrência observados nas respostas dos falantes não são compatíveis com uma teoria que pressupõe a existência de regras abstratas para gerar itens regulares e a utilização da memória associativa para os irregulares. Já que houve efeitos de freqüência de tipo e de ocorrência até mesmo em itens regulares, os resultados desafiam os pilares da Teoria de Palavras e Regras. De forma análoga, a percepção de que os falantes, em vez de usarem regras, agem por generalizações locais para formar os plurais das palavras, nos faz supor que a unidade de armazenamento no léxico mental é a palavra inteira, em vez de radicais e afixos. A partir das similaridades fonológicas e semânticas entre os itens já estocados no léxico mental do falante, redes de generalizações morfológicas emergem e novos itens são categorizados.

Ao encontrar um item léxico, o falante, na tarefa de encaixá-lo em seu espaço mental, age comparativamente, buscando possíveis similaridades entre a nova palavra e as memórias mais antigas em seu léxico mental. Havendo semelhanças, o novo item é classificado como pertencente a uma determinada classe. Dentro dessa perspectiva, esse item recém-estocado será flexionado da mesma forma que os demais membros de sua rede. Assim, esquemas locais emergem de similaridades entre as palavras armazenadas no léxico mental e fazem com que tipos pouco freqüentes na língua possam sobreviver por causa das semelhanças partilhadas entre os itens formadores de uma rede. No entanto, dentro dessas redes, a baixa freqüência de ocorrência dos itens léxicos faz com que eles acabem por sofrer pressões analógicas, conformando-se a paradigmas mais freqüentes na língua. Assim, através dessa interação dinâmica entre freqüência de tipo e de ocorrência, as representações mentais mudam e novos padrões emergem a partir da forma como a língua é usada pelos falantes.

No próximo capítulo, apresentamos as conclusões finais alcançadas com esta pesquisa e apontamos possíveis desdobramentos para este trabalho.

CAPÍTULO 7

Conclusões

Através da análise dos plurais terminados em *–ão*, *–l* e ditongo em *–u* no PB, esta tese tentou contribuir para o debate sobre como a freqüência de uso interfere nas representações mentais de informações lingüísticas.

Alinhavando os resultados de nossos experimentos com as informações etimológicas apresentadas sobre os grupos de palavras em *–ão*, *–l* e ditongo em *–u*, observamos que questões diacrônicas aliadas a efeitos de freqüência podem responder pelas variações de plurais atestadas.

No caso da pluralização em *–ão*, a análise do latim e do português arcaico corroborou a hipótese de que as generalizações em direção a *–ões* têm raízes etimológicas (já que o PB recebeu mais itens em *–ões* que *–ãos* e *–ães*) e ocorrem há alguns séculos, talvez por isso a taxa de migração de *–ãos* e *–ães* para *–ões* seja tão alta (38,9% e 14,9%, conforme Tabela 23). Assim, verificamos que, no grupo terminado em *–ão*, no século XIV, as formas singulares começaram a se fundir em *–ão*, o que levou à perda do correlato etimológico por falta de diferenciação fonética no singular. Sendo a forma de plural em *–ões* a mais freqüente (desde o português arcaico) para as palavras terminadas em *–ão*, o que se observa atualmente nesse grupo de itens é a continuação da tendência diacrônica de adotar o plural em *–ões* para todas as palavras em *–ão* no singular. Após o nivelamento fonológico das formas de singular, não houve restrições para que *–ões* (altamente freqüente) se expandisse nessa classe e tomasse o lugar dos morfemas plurais em *–ãos* e *–ães*. Esses foram preservados em apenas alguns grupos específicos, conforme vimos anteriormente.

Quanto aos itens terminados em *–l* e ditongo em *–u*, observa-se que a variação de plurais decorre da falta de distinção fonética entre *–l* e *–u* em fim de sílaba. No entanto, esse fenômeno não é, no PB, tão antigo quanto a fusão dos singulares *–ano*, *–ane* e *–one* (que se deu entre os séculos XIV e XV), por isso as generalizações de plural, nos dois grupos, ocorrem em taxas menores que no grupo em *–ão*: 8,1% para palavras terminadas em *–l* e

19,7% para itens terminados em ditongo em *–u* (Cf. Tabelas 34 e 45). As migrações registradas nesses grupos de plurais indicam que o tipo em *–is* (mais freqüente que *–us* no plural) está abarcando itens da classe em ditongo em *–u*. Já que tanto as palavras terminadas em *–l* quanto as em *–u* apresentam formas singulares em [w], está ocorrendo uma convergência fonológica entre as duas classes. Sendo assim, não havendo distinção fonética entre *–l* e *–u* em fim de sílaba, não há, também, barreiras para que a pluralização mais freqüente (em *–is*, em vez de *–us*) se aplique a itens que terminam em [w] no singular. A fusão de itens terminados em *–l* e ditongo em *–u* em uma só classe (formada a partir de similaridades fonológicas) é mais recente que a junção da classe de *–ão* no singular, portanto não podemos afirmar, precisamente, se a taxa de migração de ditongo em *–u* em direção a *–l* será, futuramente, tão alta quanto a de *–ãos* e *–ães* para *–ões*. No entanto, a julgar pelo exemplo que ocorreu com a classe de *–ão*, pode ser que estejamos assistindo, sincronicamente, a uma reordenação da morfofonologia nominal desencadeada por efeitos de freqüência (de tipo e de ocorrência) a partir da identidade fonológica entre itens léxicos específicos.

Estamos, portanto, lidando com dois grupos distintos de palavras (plurais em *–ão*, de um lado, e plurais em *–l* e ditongo em *–u*, de outro), mas que apresentam generalizações motivadas por um princípio análogo: a fusão de classes, decorrente da identidade fonológica no singular, com conseqüências para os processos morfofonológicos. Através de nossos experimentos, observamos que, niveladas as distinções fonológicas entre as formas singulares (em *–ão*, de um lado, e [w], de outro, considerando itens em *–l* e ditongo em *–u* em uma só classe), formou-se um ambiente favorável para que efeitos de freqüência pudessem ocorrer.

Nesse sentido, conforme o Modelo de Redes e o de Exemplares postulam, a interação entre freqüência de tipo e de ocorrência é, a um só tempo, conseqüência do uso da língua e desencadeador de variações nas estruturas lingüísticas. Assim, no caso dos plurais sob análise, observamos que palavras de alta freqüência de ocorrência, ainda que pertençam a padrões paradigmáticos infreqüentes, conseguem manter sua irregularidade, porque sua repetição gera representação lexical autônoma. Essas palavras, portanto, podem se desligar de sua rede e ser acessadas de forma independente, sem que seja necessário recorrer à sua classe; por sua vez, palavras de baixa freqüência de ocorrência não têm autonomia lexical e precisam ser ativadas através de sua classe. Sendo assim, se essas palavras (pouco freqüentes) pertencerem a um tipo freqüente na língua, podem se manter

imunes a mudanças analógicas; no entanto, se essas mesmas palavras pertencerem a tipos infreqüentes, provavelmente sofrerão efeito da analogia e passarão a ser flexionadas a partir de uma classe mais recorrente. É claro que, para esse processo ser desencadeado, é necessário que haja similaridades fonológicas entre classes de itens. Nos casos sob análise nesta tese, observamos que são justamente essas semelhanças que tornam possível a reorganização dos processos morfofonológicos de pluralização do PB.

Se a diferenciação fonológica entre esses grupos de palavras persistisse, provavelmente não estaria ocorrendo, atualmente, esse processo de reformulação morfológica desencadeado por efeitos de freqüência. Essa fusão fonológica das classes e seus efeitos para a morfologia dos grupos envolvidos comprovam que, conforme postulam o Modelo de Redes e o de Exemplares, não existe separação clara entre léxico e gramática, por isso diferentes níveis gramaticais (fonologia, morfologia, sintaxe) interferem mutuamente uns nos outros. Ao contrário disso, a Teoria de Palavras e Regras, que defende a separação entre léxico e gramática e define módulos gramaticais com atuações específicas (fonologia é a interface entre língua e pensamento, morfologia é o conjunto de regras para formar palavras e sintaxe é um "time" de regras para combinar palavras e sentenças[101]) não poderia justificar por que o componente fonológico da língua estaria acarretando conseqüências na morfologia de número dessas classes de plural.

Observamos que os resultados de nossos experimentos são compatíveis com um modelo de léxico mental remodelado pelo uso e que comporta informações probabilísticas acerca dos padrões disponíveis no uso diário da língua. As respostas dos falantes aos experimentos foram muito próximas das estatísticas reais da língua. Isso sugere que o uso das palavras no cotidiano afeta as representações mentais, por isso os indivíduos têm uma percepção probabilística sobre as estruturas lingüísticas e pluralizam as palavras levando em conta não apenas o segmento final (a receber a marca formal de plural), mas o item léxico como um todo e as estatísticas sobre os segmentos possíveis para sua pluralização. Assim, concluímos que o léxico é capaz de armazenar palavras inteiras, sejam elas primitivas ou já flexionadas, independentemente de sua regularidade. Como o Modelo de Redes e o de Exemplares conferem ao uso da língua papel constante na reorganização do léxico, o que define a estocagem da palavra é sua freqüência de ocorrência, não sua regularidade. Um modelo baseado em regras abstratas não poderia explicar por que houve efeitos de freqüência de ocorrência em palavras regulares e por que classes irregulares

[101] Cf. PINKER, 1999, p. 22.

(como a pluralização em *–ões*) interferiram em grupos de flexão regular (como o plural em *–ãos*).

A sobrevivência de padrões pouco freqüentes em grupos específicos, como *–ãos* e *–ães* em palavras monossílabas ou paroxítonas, sugere que o léxico mental estabelece conexões entre os itens listados. Propriedades morfológicas emergem dessas redes e fazem com que classes de palavras sejam mantidas porque a conexão entre o item primitivo e o flexionado é fortalecida a cada vez que um desses exemplares é acessado. Por outro lado, a emergência de generalizações a partir de protótipos (como o fato de que palavras terminadas em *–l* são tipicamente paroxítonas e palavras em ditongo em *–u* são oxítonas) demonstra a existência de nuvens de exemplares categorizados a partir de um membro central, que concentra a maior quantidade de características de sua classe. A partir das similaridades com esse item prototípico, os demais membros são organizados e generalizações emergem, fazendo com que padrões lexicais se mantenham.

Observamos, portanto, que a dinamicidade demonstrada pelos grupos de plurais sob análise nesta tese não pode ser acomodada em um modelo de léxico mental que postula regras para itens regulares e memorização para os irregulares, restringindo efeitos de freqüência apenas aos irregulares. Em vez disso, esse dinamismo das estruturas lingüísticas é melhor analisado por um modelo de léxico que assume o caráter múltiplo e redundante das representações mentais e que considera efeitos de freqüência como determinantes e determinados pelo uso concreto da linguagem em situações reais de comunicação. As estruturas e as relações sociais são dinâmicas, assim como o são, também, as línguas que servem aos propósitos culturais das sociedades. Conseqüentemente, as representações mentais devem preservar essa dinamicidade, renovando-se e modificando as estruturas gramaticais a partir de padrões de freqüência decorrentes do uso. Assim, os fenômenos morfofonológicos de pluralização do PB registrados nesta tese demonstram que as representações lingüísticas refletem as peculiaridades do uso da língua e que mecanismos de repetição, automatização e generalização produzem variações lingüísticas diacronicamente, cujos efeitos podem ser percebidos no estado sincrônico das línguas. Como Bybee (2003) afirma, os verdadeiros universais lingüísticos são esses mecanismos de mudança que impulsionam a constante criação e recriação da gramática.

Conforme mencionamos anteriormente, esta tese propôs uma análise inovadora da morfologia de grupos específicos de plurais do PB (em *–ão*, *–l* e ditongo em *–u*) à luz das teorias de representação mental. Tentamos analisar possíveis variações nessas

classes como reflexo da reorganização do sistema de pluralização do PB em função do uso das estruturas lingüísticas no cotidiano do falante. Como vimos ao longo desta tese, efeitos de freqüência afetam as representações mentais de informações lingüísticas e estas, por sua vez, são responsáveis pela remodelação diacrônica da estrutura lingüística. Assim, a freqüência é, a um só tempo, causa e conseqüência das variações lingüísticas. Esperamos que a análise aqui proposta tenha sido capaz de demonstrar que o Modelo de Redes e o de Exemplares – e não teorias que postulam a existência de regras abstratas – apresentam concepções teóricas mais compatíveis com o quadro de pluralização observado.

Esta tese não tem como objetivo esgotar o tema que, em decorrência da amplitude e dinamicidade das questões relativas ao papel da freqüência nas representações mentais de informações lingüísticas, sempre carece de análises mais aprofundadas. Em função disso, possíveis desdobramentos para esse assunto seriam:

- ✓ Elaboração de experimentos com palavras inventadas (pertencentes aos grupos de plural em *–ão*, *–l* e ditongo em *–u*) a fim de verificar a possível generalização de padrões de tipos mais freqüentes e as restrições fonotáticas impostas à aplicação desses paradigmas;
- ✓ Aplicação de experimentos (com palavras reais ou inventadas) a falantes não--nativos do PB, com o objetivo de observar se o tamanho reduzido do léxico mental (sobre o PB) desses informantes interferiria nas possíveis generalizações decorrentes de efeitos de freqüência;
- ✓ Acompanhamento da aquisição das marcas de pluralização por crianças nativas do PB, a fim de verificar se elas apresentam padrões de aquisição compatíveis com a aprendizagem em formato de "U"[102] e se há faixas etárias mais propensas às generalizações de plural.

Essas possibilidades de pesquisa acrescentariam informações complementares e igualmente relevantes para o debate sobre os efeitos de freqüência nas representações mentais e sobre como o léxico mental categoriza, armazena e acessa itens lingüísticos. Fica, portanto, o convite promissor para que essas investigações, em um tempo futuro, possam ser realizadas.

---

[102] Conforme vimos no Capítulo 4, "Quadro Teórico", a aprendizagem em formato de "U" refere-se aos seguintes estágios de flexão: 1) Domínio de itens irregulares; 2) Regularização de irregulares; 2) Coexistência das formas regulares e irregulares.

REFERÊNCIAS

ABAURRE, Maria Bernadete Marques. *Nasality in Portuguese:* a critical consideration of a proposed analysis for word-final diphthongs. 1973. 28 f. University of New Mexico, Albuquerque, 1973.

ALMEIDA, Napoleão Mendes de. *Gramática metódica da língua portuguesa*. 41. ed. São Paulo: Saraiva, 1997.

AMARAL, Amadeu. *O dialecto caipira*. São Paulo: Casa Editora "O Livro", 1920.

BERENT, Iris; PINKER, Steven; SHIMRON, Joseph. The nature of regularity and irregularity: evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research*, New York, v. 31, n. 5, p. 459-502, Set. 2002.

BRAGA, Maria Luiza; SCHERRE, Maria Marta Pereira. A concordância de número no SN na área urbana do Rio de Janeiro. In: ENCONTRO NACIONAL DE LINGÜÍSTICA, 1º, 1976, Rio de Janeiro. *Anais*... Rio de Janeiro, [S.n], 1976. p. 464-77 *apud* SCHERRE, Maria Marta Pereira. *Reanálise da concordância nominal em português*. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

BROWMAN, Catherine P.; GOLDSTEIN, Louis M. Articulatory phonology: An overview. *Phonetica*, [S.l.], n. 49 (3-4), p. 155-80, 1992.

BYBEE, Joan. From usage to grammar: the mind's response to repetition. *Language*, Washington, n. 82(4), p. 529-551, 2006.

_____. Mechanisms of change in grammaticization: the role of frequency. In: JANDA, Richard; JOSEPH, Brian (Eds.). *Handbook of historical linguistics*. Oxford: Blackwell, 2002a. p. 602-623.

_____. Mechanisms of change as universals of language. 2003, [S.l. : s.n.].

_____. *Morphology:* a study of the relation between meaning and form. Philadelphia: John Benjamins, 1985.

_____. *Phonology and language use.* Cambridge: Cambridge University Press, 2001. (Cambridge Studies in Linguistics, 94).

_____. Productivity, regularity and fusion: how language use affects the lexicon. In: SINGH, Rajendra; DESROCHERS, Richard (Eds.). *Trubetzkoy's orphan:* Proceedings of the Montréal Roundtable "Morphonology: Contemporary responses". Amsterdam/Philadelphia: John Benjamins Publishing Company, 1994, p. 249-294.

_____. Regular morphology and the lexicon. *Language and Cognitive Processes*, Cambridge, n. 10, p. 425-455, 1995.

_____. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, Cambridge, n. 14, p. 261-290, 2002b.

BYBEE, Joan; NEWMAN, Jean. Are stem changes as natural as affixes? *Linguistics*, Mouton de Gruyter, Berlin/New York, n. 33, p. 633-654, 1995.

BYBEE, Joan; MODER, Carol. Morphological classes as natural categories. *Language*, Washington, n. 59, p. 251-270, 1983.

BYBEE, Joan; SLOBIN, Dan. 1982. Rules and schemas in the development and use of the English past tense. *Language*, Washington, n. 58, p. 265-289, 1982.

BYBEE, Joan; SCHEIBMAN, Joanne. The effect of usage on degrees of constituency: the reduction of *don't* in English. *Linguistics*, Mouton de Gruyter, Berlin/New York, n. 37, p. 575-596, 1999.

CAMARA JÚNIOR, Joaquim Mattoso. *Estrutura da língua portuguesa*. 20. ed. Petrópolis: Vozes, 1991.

_____. *The Portuguese language*. Tradução de Anthony Julius Naro. Chicago: The University of Chicago Press, 1972.

CAMPOS, Odette G. L. Altmann de Souza. A origem e a flexão dos nomes terminados em –ão na língua portuguesa. In: *Estudos de filologia e lingüística:* em homenagem a Isaac Nicolau Salum. São Paulo: T. A. Queiroz: Editora da Universidade de São Paulo, 1981. p. 147-157.

CHAMBERS, John Kenneth. *Sociolinguistic theory:* Linguistic variation and its social significance. Oxford/Cambridge: Blackwell, 1995.

CHOMSKY, Noam; HALLE, Morris. *The sound pattern of English*. New York: Harper and Row, 1968 *apud* BYBEE, Joan. *Morphology:* a study of the relation between meaning and form. Philadelphia: John Benjamins, 1985.

CHOMSKY, Noam; LASNIK, Howard. The theory of principles and parameters. In: CHOMSKY, Noam (Ed.). *The minimalist program*. Cambridge, MA: MIT Press, 1995. p. 13-128 *apud* PIERREHUMBERT, Janet. Exemplar dynamics: word frequency, lenition and contrast. In: BYBEE, Joan; HOOPER, Paul (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001b. p. 137-57.

CHRISTIANSEN, Morten; CHATER, Nick. Connectionist psycholinguistics in perspective. In: M.H. CHRISTIANSEN, Morten; CHATER, Nick. (Eds.). *Connectionist psycholinguistics*. Westport, CT: Ablex, 2000. p. 19-75.

CIELO, Carla Aparecida. A flexibilidade do paradigma conexionista. *Letras de Hoje,* Porto Alegre, v. 33. n. 2, p. 43-50, Junho 1998.

Corpus ASPA (Avaliação Sonora do Português Atual). Disponível em: <http://www.projetoaspa.org/>. Acesso em diferentes fases da redação desta tese.

Corpus LAEL (Lingüística Aplicada e Estudos da Linguagem). Disponível em: <http://lael.pucsp.br/corpora/index.htm>. Acesso em diferentes fases da redação desta tese.

Corpus NILC/São Carlos. Disponível em: <www.linguateca.pt/ACDC/>. Acesso em diferentes fases da redação desta tese.

COUTINHO, Ismael de Lima. *Pontos de gramática histórica*. 7. ed. Rio de Janeiro: Livraria Acadêmica, 1981.

CRISTÓFARO-SILVA, Thaïs. *Fonética e fonologia do português:* roteiro de estudos e guia de exercícios. 6. ed. São Paulo: Editora Contexto, 2002.

CRISTÓFARO-SILVA, Thaïs; GOMES, Christina Abreu; GUIMARÃES, Daniela; HUBACK, Ana Paula da Silva. The acquisition of irregular plurals in Brazilian Portuguese. In: ASSOCIATION FOR THE STUDY OF CHILD LANGUAGE, 2005, Berlin. Association for the Study of Child Language Abstract Booklet, 2005. p. 168-169.

CUNHA, Antônio Geraldo da. *Dicionário etimológico nova fronteira da língua portuguesa*. Rio de Janeiro: Nova Fronteira, 1982.

CUNHA, Celso; CINTRA, Lindley. *Nova gramática do português contemporâneo*. 2. ed. Lisboa: Edições João Sá da Costa, 1985.

DAUGHERTY, Kim; SEIDENBERG, Mark. Beyond rules and exceptions: a connectionist approach to inflectional morphology. In: LIMA, Susan; CORRIGAN, Roberta; IVERSON, Gregory (Eds.). *The reality of linguistic rules*. Amsterdam/Philadelphia: John Benjamins, 1994, p. 353-388.

Dicionário Houaiss on-line. Disponível em: <http://houaiss.uol.com.br/busca.jhtm>. Acesso em: 24 de dezembro de 2004.

Elman, Jeffrey L. An alternative view of the mental lexicon. *Trends in Cognitive Science,* Oxford, v. 7, 301-306, 2004.

_____. Connectionism, artificial life, and dynamical systems: new approaches to old questions. In: BECHTEL, William; GRAHAM, George (Eds.). *A Companion to Cognitive Science.* Oxford: Basil Blackwood. 1998a. [s.p.]

_____. Finding structure in time. *Cognitive science*, [S.l.], n. 14, p. 179-211, 1990.

_____. Generalization, simple recurrent networks, and the emergence of structure. In: GERNSBACHER, Morton Ann; DERRY, Sharon (Eds.). *Proceedings of the 20th Annual Conference of the Cognitive Science Society.* Mahway: Lawrence Erlbaum Associates, 1998b. p. 1-17.

FERNANDES, Xavier. *Questões de língua pátria.* Lisboa: Edição da Revista 'Ocidente', 1947. v. II.

FERREIRA, Aurélio Buarque de Holanda. *Novo dicionário Aurélio da língua portuguesa.* 3. ed. Curitiba: Positivo, 2004.

FIDELHOLTZ, James L. Word frequency and vowel reduction in English. In: REGIONAL MEETING CHIGAGO LINGUISTICS SOCIETY, 7, 1975, Chicago. *Papers from...* Chicago: Chicago Linguistic Society, 1975. p. 200-213.

FONTES MARTINS, Raquel Márcia. *A organização do componente fonológico e o comportamento do indivíduo.* 2007. Tese (Doutorado em Lingüística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007. A ser publicada.

_____. *O cancelamento das líquidas /ʎ/ e /ɾ/ intervocálicas no português de Belo Horizonte.* 2001. Dissertação (Mestrado em Lingüística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2001.

FRANÇA, Júnia Lessa; VASCONCELLOS, Ana Cristina. *Manual para normalização de publicações técnico-científicas.* 7. ed. Belo Horizonte: Editora UFMG, 2004.

FRANCIS, W. Nelson; KUČERA, Henry. Frequency analysis of English usage. Boston, MA: Houghton-Mifflin, 1982 *apud* MacWHINNEY, Brian.; LEINBACH, Jared. Implementations are not conceptualizations: revising the verb learning model. *Cognition*, Amsterdam/New York/Oxford/Shannon/Tokyo, v. 40, p. 121-157, 1991.

FRISCH, Stefan A. *et al.* Emergent phonotactic generalizations in English and Arabic. In: BYBEE, Joan; HOOPER, Paul (Eds.). *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins, 2001b. p. 159-179.

GOLDINGER, Stephen D. Words and voices: traces in spoken word identification and recognition memory. *Journal of Experimental Psychology:* Learning, Memory, and Cognition, Washington, v. 22 (5), p. 1166-1183, 1996.

GUERIOS, Rosario Farani Mansur. *Pontos de gramática histórica portuguesa.* 2. ed. São Paulo: Livraria Acadêmica, 1942.

GUY, Gregory R. *Linguistic variation in Brazilian Portuguese:* aspects of the phonology, syntax, and language history. 1981. 391p. Tese (Doutorado em Lingüística) – University of Pennsylvania, Philadelphia,1981.

HARE, Mary; ELMAN, Jeffrey. Learning and morphological change. *Cognition,* [S.l.], v. 56, p. 61-98, 1995.

HAY, Jennifer. Lexical frequency in morphology: is everything relative? *Linguistics,* Mouton de Gruyter, Berlin/New York, n. 39 (6), p. 1041-1070, 2001.

HAY, Jennifer; BAAYEN, Harald. Parsing and productivity. In: BOOIJ, Geert; VAN MARLE, JAAP (Eds.). *Yearbook of Morphology*. Amsterdam: 2001. p. 203-235.

HOOPER, Joan Bybee. Word frequency in lexical diffusion and the source of morphophonological change. In: CHRISTIE, William. *Current progress in historical linguistics*. Amsterdam: North Holland, 1976. p. 96-105.

HOPPER, Paul. Emergent grammar. In: TOMASELLO, Michael (Ed.). *The new psychology of language:* cognitive and functional approaches to language structure. Mahwah/New Jersey/ London: Lawrence Erlbaum Associates, Inc., 1998. p. 155-175.

HSIEH, Hsin-I. Lexical diffusion: evidence from child language acquisition. In: WANG, William S.-Y. (Ed.). *The lexicon in phonological change*. The Hague: Mouton , 1977. p. 133-147.

HUME, Elizabeth; JOHNSON, Keith. A model of the interplay of speech perception and phonology. In: HUME, Elizabeth; JOHNSON, Keith (Eds.). *The Role of Speech Perception in Phonology*, London: Academic Press, 2001. p. 1-17.

JOHNSON, Keith. Speech perception without speaker normalization: an exemplar model. In: JOHNSON, Keith; MULLENNIX, John (Eds.). *Talker Variability in Speech Processing*. San Diego: Academic Press, 1997. p. 145-165.

JOHNSON, Keith; MULLENIX, John (Eds.). *Talker Variability in Speech Processing*. San Diego: Academic Press, 1997.

JUCÁ FILHO, Cândido. *Língua nacional:* as diferenciações entre o português de Portugal e o do Brasil autorizam a existência de um ramo dialetal do português peninsular? Rio de Janeiro: Livraria Francisco Alves, 1937.

KAPATSINSKI, Vselvolod M. *Productivity of Russian stem extensions:* evidence for and a formalization of network theory. 2005. 180 p. Dissertação (Mestrado em Lingüística) – University of New Mexico, Albuquerque, 2005.

KEMMER, Suzanne; BARLOW, Michael. Introduction: A usage-based conception of language. In: KEMMER, Suzanne; BARLOW, Michael (Eds.). *Usage-based models of*

language. Stanford: Center for the study of language and information, 2000. p. vii-xxviii.

KÖPCKE, Klaus-Michael. Schemas in German plural formation. *Lingua*, [S.l.], v. 74, p. 303-335, 1988 *apud* BYBEE, Joan. Productivity, regularity and fusion: how language use affects the lexicon. In: SINGH, Rajendra; DESROCHERS, Richard (Eds.). *Trubetzkoy's orphan:* Proceedings of the Montréal Roundtable "Morphonology: Contemporary responses". Amsterdam/Philadelphia: John Benjamins Publishing Company, 1994, p. 249-294.

KUčERA, Henry; FRANCIS, W. Nelson. Computational analysis of present-day American English. Providence, RI: Brown University Press, 1967 *apud* RUMELHART, David; McCLELLAND, James. On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In: McCLELLAND, James; RUMELHART, David *et al.* (Orgs.). *Parallel distributed processing:* explorations in the microstructure of cognition. Cambridge: MIT Press, 1986. p. 216-71.

KURY, Adriano da Gama; BUENO, José Renato; OLIVEIRA, Ubaldo Luiz de. *Gramática objetiva da língua portuguesa:* sintaxe. 2. ed. Rio de Janeiro: Editora Rio, 1977.

LABOV, William. *Principles of linguistic change:* internal factors. Oxford: Blackwell, 1994.

_____. *Principles of linguistic change:* social factors. Oxford: Blackwell, 2001. v. 2.

_____. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

_____. *The social stratification of English in New York*. Washington, D.C.: Center for Applied Linguistics, 1966a *apud* LABOV, William. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LACERDA, Francisco. Distributed memory representations generate the perceptual-magnet effect. *Journal of the Acoustical Society of America*. In press *apud* PIERREHUMBERT, Janet. Exemplar dynamics: word frequency, lenition and

contrast. In: BYBEE, Joan; HOOPER, Paul (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001b. p. 137-57.

LANGACKER, Ronald. Foundations of cognitive grammar. Volume 1 – Theoretical prerequisites. Stanford: Stanford University Press, 1987.

LESLAU, Wolf. Frequency and change in the Ethiopian languages. *Word*, New York, v. 25, p. 180-189, 1969.

MacWHINNEY, Brian; LEINBACH, Jared. Implementations are not conceptualizations: revising the verb learning model. *Cognition*, [S.l.], v. 40, p. 121-157, 1991.

MARCUS, Gary F. *et al*. German inflection: the exception that proves the rule. *Cognitive Psychology,* [S.l.], v. 29, p. 189-256, 1995, *apud* BYBEE, Joan. Productivity, regularity and fusion: how language use affects the lexicon. In: SINGH, Rajendra; DESROCHERS, Richard (Eds.). *Trubetzkoy's orphan:* Proceedings of the Montréal Roundtable "Morphonology: Contemporary responses". Amsterdam/Philadelphia: John Benjamins Publishing Company, 1994, p. 249-294.

MARROQUIM, Mario. *A língua do Nordeste (Alagôas e Pernambuco)*. 2. ed. São Paulo/Rio de Janeiro/Recife/Bahia/Pará/Pôrto Alegre: Companhia Editora Nacional, 1945, v. 25. (Biblioteca Pedagógica Brasileira, Série 5ª).

MATTOS E SILVA, Rosa Virgínia. *Estruturas trecentistas:* elementos para uma gramática do português arcaico. Rio de Janeiro: Imprensa Nacional, [19--]. (Série Universitária).

McCLELLAND, James; PATTERSON, Karalyn. Rules or connections in past-tense inflections: what does the evidence rule out? *TRENDS in Cognitive Sciences*, Oxford, v. 6, n. 11, p. 465-472, 2002b.

McCLELLAND, James; PATTERSON, Karalyn. 'Words *or* rules' cannot exploit the regularity in exceptions. *TRENDS in Cognitive Sciences*, Oxford, v. 6, n. 11, p. 464-465, 2002a.

Michaelis Moderno Dicionário do Português Brasileiro. Disponível em: <http://www2.uol.com.br/michaelis/>. Acesso em diferentes fases da redação desta tese.

MIRA MATEUS, Maria Helena *et al. Gramática da língua portuguesa.* 6. ed. Lisboa: Caminho, 2003. (Colecção Universitária. Série Lingüística).

MÜLLER, Ana. Generic sentences with indefinite and bare subjects in Brazilian Portuguese. In: ANDERSSEN, Jan; MENÉNDEZ-BENITO, Paula; WERLE, Adam (Eds.). *The Proceedings of SULA 2*. Vancouver, [s.n.], 2003. p. 71-86.

NARO, Anthony Julius. The social and structural dimensions of a syntactic change. *Language*, Washington, v. 57, n. 1, p. 63-98, Mar. 1981.

NARO, Anthony Julius; VOTRE, Sebastião José. *SWAVA*: Sistema SWAMINC/VARBRUL (Manual do Usuário). UFRJ, 1980. Mimeografado *apud* SCHERRE, Maria Marta Pereira. *Reanálise da concordância nominal em português*. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

NARO, Anthony Julius; SCHERRE, Marta Maria Pereira. Variable concord in Portuguese: the situation in Brazil and Portugal. In: McWHORTER, John (Ed.). *Language change and language contact in pidgins and creoles*. Amsterdam/Philadelphia: John Benjamins, 2000. v. 21, p. 235-255.

NASCENTES, Antenor. *Dicionário etimológico da língua portuguesa*. Rio de Janeiro: Livraria Francisco Alves, 1932. v. I.

_____. *O linguajar carioca*. Rio de Janeiro: Organização Simões, 1953.

NÓBREGA, Vandick L. da; MEDEIROS, Walter. *O idioma do Brasil*: antologia, gramática, exercícios. São Paulo: Companhia Editora Nacional, 1965. v. 1.

NUNES, José Joaquim. *Compêndio de gramática histórica portuguesa (fonética e morfologia)*. 3. ed. Lisboa: Livraria Clássica Editora, 1945.

_____. *Digressões lexicológicas*. Lisboa: Livraria Clássica Editora, 1928.

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

_____. Reanalisando 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.

_____. The neogrammarian controversy revisited. *International Journal of the Sociology of Language,* Berlin, v. 89, p. 93-105, 1991.

PAGLIUCA, William; MOWREY, Richard. Articulatory evolution. In: RAMAT, Anna Giacalone; CARRUBA, Onofrio; BERNINI, Giuliano. *Papers from the 7th International Conference on Historical Linguistics*. Amsterdam: Benjamins, 1987. p. 459-72.

PAIVA, Maria da Conceição de. A variável gênero/sexo. In: MOLLICA, Maria Cecilia; BRAGA, Maria Luiza (Orgs.). *Introdução à sociolingüística:* o tratamento da variação. São Paulo: Contexto, 2003. p. 33-42.

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

PARREIRA, Manuela; PINTO, J. Manuel de Castro. *Prontuário ortográfico moderno*. Lisboa: Edições Asa, 1985.

PETERSON, Gordon E.; BARNEY, Harold L. Control methods used in a study of the vowels. *Journal of the acoustic society of America*, [S.l.], v. 24, p. 175-184, 1952 *apud* PIERREHUMBERT, Janet. Probabilistic phonology: discrimination and robustness. In: BOD, Rens; HAY, Jennifer; JANNEDY, Stefanie. (Eds.). *Probability Linguistics*. Cambridge/London: MIT Press, 2003. p. 177-228.

PHILLIPS, Betty S. Word frequency and the actuation of sound change. *Language*, Washington, v. 60, n. 2, p. 320-42, 1984.

PIERREHUMBERT, Janet. An unnatural process. Written version of paper delivered at 8th meeting on Laboratory Phonology, New Haven, 6/02. Under review for Laboratory Phonology 8, Mouton de Gruyter, 2002. In press.

_____. Exemplar dynamics: word frequency, lenition and contrast. In: BYBEE, Joan; HOOPER, Paul (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001b. p. 137-157.

_____. Probabilistic phonology: discrimination and robustness. In: BOD, Rens; HAY, Jennifer; JANNEDY, Stefanie. (Eds.). *Probability Linguistics*. Cambridge/London: MIT Press, 2003. p. 177-228.

_____. Stochastic phonology. In: *GLOT International*. Oxford, v. 5, n. 6, p. 195-297, 2001c.

_____. What people know about sounds of language. *Studies in the Linguistic Sciences,* Urbana, n. 29, p. 111-120, 2000.

_____. Word-specific phonetics. In: GUSSENHOVEN, Carlos; WARNER, Natasha (Eds.). *Laboratory Phonology VII*. Berlin: Mouton de Gruyter, 2001a. p. 1-24.

PINKER, Steven. Rules of language. *Science*, [S.l.], n. 253, p. 530-535, 1991.

_____. *Words and rules:* the ingredients of language. New York: Perennial, 1999.

PINKER, Steven; ULLMAN, Michael. Combination and structure, not gradedness, is the issue. *TRENDS in Cognitive Sciences*, Oxford, v. 6. n. 11, p. 472-74. 2002b.

PINKER, Steven; PRINCE, Alan. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, [S.l.], n. 28, p. 73-193, 1988.

PINKER, Steven; PRINCE, Alan. Regular and irregular morphology and the psychological status of rules of grammar. In: LIMA, Susan; CORRIGAN, Roberta; IVERSON, Gregory (Eds.). *The reality of linguistic rules*. Amsterdam/Philadelphia: John Benjamins, 1994. p. 321-351.

PINKER, Steven; ULLMAN, Michael. The past and future of the past tense. *TRENDS in Cognitive Sciences*, Oxford, v. 6. n. 11, p. 456-463, 2002a.

POPLACK, Shana. The notion of the plural in Puerto Rico Spanish: competing constraints on /s/ deletion. In: LABOV, William (Ed.). *Locating language in time and space*. New York: Academic, 1991. p. 55-67 *apud* SCHERRE, Marta Maria Pereira; Naro, Anthony Julius. Marking in discourse: "birds of a feather". *Language Variation and Change*, Cambridge, v. 3, p. 23-32, 1991.

PLUNKETT, Kim. O conexionismo hoje. Tradução de Márcia Zimmer e Rosângela Gabriel. *Letras de Hoje*, Porto Alegre, v. 35, n. 4, p. 109-122, Dez. 2000.

PLUNKETT, Kim; MARCHMAN, Virginia A. U-shaped learning and frequency effects in a multi-layered perceptron. *Cognition*, [S.l.], n. 39, p. 43-102, 1991.

PRASADA, Sandeep; PINKER, Steven; SNYDER, William. Some evidence that irregular forms are retrieved from memory but regular forms are rule generated. Paper presented at the *Thirty-first Annual Meeting of the Psychonomic Society*, New Orleans, [s.n.], 1990 *apud* PINKER, Steven. *Words and rules:* the ingredients of language. New York: Perennial, 1999.

Projeto Censo. Disponível em: <http://www.letras.ufrj.br/~peul/amostras.htm>. Acesso em: 25 de novembro de 2005.

ROCHA LIMA, Carlos Henrique da. *Gramática normativa da língua portuguesa*. Rio de Janeiro: F. Briguiet & Cia. Editores, 1984.

ROSCH, Eleanor; MERVIS, Carolyn. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, [S.l.], v. 7, p. 573-605, 1975 *apud* BYBEE, Joan; MODER, Carol. Morphological classes as natural categories. *Language*, Washington, n. 59, p. 251-270, 1983.

ROSCH, Eleanor. Principles of categorization. In: ROSCH, Eleanor; LLOYD, Barbara B. (Eds.). *Cognition and categorization*. Hillsdale: Erlbaum, 1978. p. 27-48. *apud* BYBEE, Joan; MODER, Carol. Morphological classes as natural categories. *Language*, Washington, n. 59, p. 251-270, 1983.

ROUSSEAU, Pascale & SANKOFF, David. Advances in variable rule methodology. In: SANKOFF, David. (Ed.). *Linguistic variation*: *models and methods*. New York: Academic Press, 1978. p. 57-68 *apud* SCHERRE, Maria Marta Pereira. *Reanálise da concordância nominal em português*. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

RUMELHART, David; McCLELLAND, James. On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In: McCLELLAND, James; RUMELHART, David *et al.* (Orgs.). *Parallel distributed processing:* explorations in the microstructure of cognition. Cambridge: MIT Press, 1986. p. 216-71.

SAID ALI, Manuel. *Gramática histórica da língua portuguesa*. 3. ed. São Paulo: Edições Melhoramentos, 1964.

SANGSTER, Catherine M. *Inter- and Intra-Speaker Variation in Liverpool English*: A Sociophonetic Study. 2002. 290 f. Tese (Doutorado) – University of Oxford, Oxford, 2002.

SANKOFF, David. *Varbrul Version 2*. Centre de Réchérches Mathématiques, Université de Montréal, 1975. Mimeografado *apud* SCHERRE, Maria Marta Pereira. *Reanálise da*

concordância nominal em português. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

_____. *Varbrul 2S*. 1979. Não publicado *apud* SCHERRE, Maria Marta Pereira. *Reanálise da concordância nominal em português*. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

SCHERRE, Maria Marta Pereira. *A regra de concordância de número no sintagma nominal em português*. 1978. 158 f. Dissertação (Mestrado em Lingüística) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1978 *apud* NARO, Anthony Julius. The social and structural dimensions of a syntactic change. *Language*, Washington, v. 57, n. 1, p. 63-98, Mar. 1981.

_____. Aspectos da concordância de número no português do Brasil. *Revista Internacional de Língua Portuguesa (RILP):* Norma e Variação do Português. Lisboa, n. 12, p. 37-49, Dez. 1994.

_____. *Doa-se lindos filhotes de poodle:* variação lingüística, mídia e preconceito. São Paulo: Parábola, 2005.

_____. Major linguistic patterns in noun phrase agreement in Brazilian Portuguese. In: Cinquanta'anni di recerche linguistiche: problemi, resultati e prospettive per il terzo millennio. Atti del IX convegno internazionale di linguisti – tenuto a Milano nei giorni 8-18 ottobre 1998. Alessandria: Edizioni dell'Orso S.r.l. (ISBN 88-7694-526-1) Sodalizio Glottologico Milanese – Instituto Lombardo – Academia di Scienze e Leterre. p. 461-473, 2001a.

_____. Phrase-level parallelism effect on noun phrase number agreement. *Language Variation and Change*, Cambridge, n. 13, p. 91–107, 2001b.

_____. *Reanálise da concordância nominal em português*. 1988. 554 f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

_____. Sobre a influência de variáveis sociais na concordância nominal. In: OLIVEIRA e SILVA, Giselle Machline de; SCHERRE, Maria Marta Pereira. *Padrões sociolingüísticos:* análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro. 2. ed. Rio de Janeiro: Tempo Brasileiro, 1998. p. 239-264.

SCHERRE, Marta Maria Pereira; NARO, Anthony Julius. Marking in discourse: "birds of a feather". *Language Variation and Change*, Cambridge, v. 3, p. 23-32, 1991.

SCHERRE, Marta Maria Pereira; NARO, Anthony Julius. Análise quantitativa e tópicos de interpretação do Varbrul. In: MOLLICA, Maria Cecilia; BRAGA, Maria Luiza (Orgs.). *Introdução à sociolingüística:* o tratamento da variação. São Paulo: Contexto, 2005. p. 146-177.

_____. The serial effect on internal and external variables. *Language Variation and Change*, Cambridge, v. 4, p. 1-13, 1992.

SCHUCHARDT, Hugo. On sound laws: against the neogrammarians. In: *Schuchardt, the neogrammarians, and the transformational theory of phonological change*. Frankfurt: Athenaeum, 1972 [1885]. p. 39-72.

SEIDENBERG, Mark. Connectionism without tears. In: DAVIS, Steven (Ed.). *Connectionism:* Theory and practice. Oxford: Oxford University Press, 1992 *apud* DAUGHERTY, Kim; SEIDENBERG, Mark. Beyond rules and exceptions: a connectionist approach to inflectional morphology. In: LIMA, Susan; CORRIGAN, Roberta; IVERSON, Gregory (Eds.). *The reality of linguistic rules*. Amsterdam/Philadelphia: John Benjamins, 1994. p. 353-388.

_____. Language acquisition and use: Learning and applying probabilistic constraints. *Science*, [S.l.], v. 275, p. 1599-1603, Mar. 1997.

SEQUEIRA, Francisco Martins de. *Aspectos do português arcaico*. Lisboa: Livraria Popular, 1943.

SILVA, Marinalva Freire da. *Plural dos nomes em –ão na língua portuguesa:* uma abordagem filológica. João Pessoa: CCHLA/Idéia, 1993.

THOMAS, Earl W. *The syntax of spoken Brazilian Portuguese.* Nashville: Vanderbilt University Press, 1969.

TREIMAN, Rebecca *et al.* English speakers' sensitivity to phonotactic patterns. In: BROE, Michael; PIERREHUMBERT, Janet. *Papers in laboratory phonology V:* acquisition and the lexicon. Cambridge: Cambridge University Press, 2000, p. 269-282 *apud* PIERREHUMBERT, Janet. What people know about sounds of language. *Studies in the Linguistic Sciences,* Urbana, Illinois, n. 29, p. 111-120, 2000.

VENNEMANN, Theo. Words and syllables in natural generative phonology. In: BRUCK, Anthony; FOX, Robert; LAGALY, Michael (Eds.). *Parasession on natural phonology.* Chicago: Chicago Linguistic Society, 1974, p. 346-374 *apud* BYBEE, Joan. *Morphology:* a study of the relation between meaning and form. Philadelphia: John Benjamins, 1985.

VOTRE, Sebastião Josué. Relevância da variável escolaridade. In: MOLLICA, Maria Cecilia; BRAGA, Maria Luiza (Orgs.). *Introdução à sociolingüística:* o tratamento da variação. São Paulo: Contexto, 2003. p. 51-57.

WANG, William S.-Y.; CHENG, Chin-Chuan. Implementation of phonological change: the Shaung-feng Chinese case. In: WANG, William S.-Y. (Ed.) *The lexicon in phonological change.* The Hague: Mouton, 1977. p. 148-158.

WICKELGREN, Wayne A. Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychology Review*, Washington, D.C. n. 86, p. 44-60, 1969 *apud* RUMELHART, David; McCLELLAND, James. On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In: McCLELLAND, James; RUMELHART, David *et al.* (Orgs.). *Parallel distributed processing:* explorations in the microstructure of cognition. Cambridge: MIT Press, 1986. p. 216-71.

WILLIAMS, Edwin B. *First Brazilian grammar*: a course in beginner's Portuguese. New York: Appleton-Century-Crofts, Inc., 1944.

WITTGENSTEIN, Ludwing. *Philosophical investigations*. New York: Macmillan, 1953 *apud* BYBEE, Joan; MODER, Carol. Morphological classes as natural categories. *Language*, Washington, n. 59, p. 251-270, 1983.

ANEXOS

Anexo A

Termo de Consentimento Livre e Esclarecido

Título do Projeto: "Modelos de Léxico Mental e Grupos de Plurais do Português Brasileiro"

ATENÇÃO: Este termo de consentimento pode conter palavras que você não entenda. Peça ao pesquisador que explique as palavras ou informações sobre partes que você não entendeu.

Solicito a participação de _____ (nome do Participante) em uma pesquisa científica realizada pela aluna Ana Paula da Silva Huback, sob orientação da Professora Doutora Thaïs Cristófaro Silva, ambas vinculadas à Faculdade de Letras da Universidade Federal de Minas Gerais. A presente pesquisa é um requisito à obtenção do título de Doutorado em Lingüística e tem como objetivo analisar se fatores sociais, como faixa etária, sexo e escolaridade, interferem na forma como as pessoas memorizam figuras e palavras. Esta pesquisa apresenta o benefício de contribuir para as investigações sobre o armazenamento de palavras na memória do falante.

Caso decida participar desta pesquisa, é importante que você leia algumas informações que seguem abaixo. A participação não é obrigatória e poderá ser interrompida a qualquer momento, caso julgue necessário. A saída do estudo não trará nenhum prejuízo na relação com as pesquisadoras ou com a instituição.

O procedimento desta pesquisa consiste em fazer comentários sobre figuras, ler algumas frases e falar o plural de algumas palavras. A fala será gravada em gravador digital e, posteriormente, os dados serão analisados. A pesquisa não apresenta nenhum risco para o indivíduo que for participar dela.

A participação na pesquisa não acarretará gasto para você, sendo totalmente gratuita. Você não será identificado(a) quando o material de registro for utilizado, seja para propósitos de publicação científica ou educativa. No entanto, algumas informações obtidas a partir da participação neste estudo não poderão ser mantidas estritamente confidenciais. Além dos estudiosos que estarão realizando a pesquisa, agências governamentais locais e o Comitê de Ética em Pesquisa da instituição onde o estudo está sendo realizado

podem precisar consultar os registros. Ao assinar este consentimento, você autoriza o acesso aos seus registros.

Você receberá uma cópia deste termo onde constam o telefone das pesquisadoras, podendo tirar suas dúvidas sobre o estudo agora ou a qualquer momento. Caso você tenha mais perguntas sobre este trabalho, favor ligar para os números abaixo.

Pesquisadoras Responsáveis:

Nome: Thaïs Cristófaro-Silva (Orientadora)
Faculdade de Letras – UFMG
Endereço: Rua Herval, 556 – Serra
Belo Horizonte – MG
CEP: 30240-010
Telefone: (31) 3281-4679

Nome: Ana Paula da Silva Huback (Aluna Doutorado)
Faculdade de Letras – UFMG
Endereço: Rua Alcides Lima, 51 – Banquete
Bom Jardim – RJ
CEP: 28662-000
Telefones: (22) 2565-1187
(22) 8126-6497

COEP UFMG
Avenida Antônio Carlos, 6.627
Unidade Administrativa II
2º Andar – Sala 2005
Pampulha
Telefones: (31) 3499-4592
(31) 3499-4027

Agradecemos antecipadamente.

Declaração de consentimento

Li ou alguém leu para mim as informações contidas neste documento antes de assinar este termo de consentimento.

Declaro que tive tempo suficiente para ler e entender as informações acima e que recebi respostas para todas as minhas dúvidas. Confirmo, também, que recebi uma cópia deste formulário de consentimento. Compreendo que sou livre para me retirar do estudo em qualquer momento, sem perda de benefícios ou qualquer outra penalidade.

Participo de livre e espontânea vontade deste estudo.

_____ (Local e Data)

_____ (Assinatura do Participante)

_____

Nome do Participante (em letra de fôrma)

Atesto que o(a) participante recebeu todas as informações necessárias, que foram fornecidas em uma linguagem adequada e compreensível e que ele/ela compreendeu essa explicação.

_____	_____
Local e Data	Assinatura do pesquisador

Material utilizado nos experimentos[103]

1. Conjunto de Figuras

Palavra esperada: plural de "mão"

Palavra esperada: plural de "avental"

Palavra esperada: plural de "gato"

Palavra esperada: plural de "caminhão"

[103] Neste Anexo, mostramos as figuras e palavras exatamente na ordem em que foram apresentadas para os informantes, considerando, inclusive, os itens distratores.

Palavra esperada: plural de “berimbau”

Palavra esperada: plural de “pulmão”

Palavra esperada: plural de “farol”

Palavra esperada: plural de “árvore”

Palavra esperada: plural de “avião”

Palavra esperada: plural de “anel”

Palavra esperada: plural de “móvel”

Palavra esperada: plural de “grão”

Palavra esperada: plural de “estrela”

Palavra esperada: plural de “pão”

Palavra esperada: plural de “cachecol”

Palavra esperada: plural de “pincel”

Palavra esperada: plural de “chapéu”

Palavra esperada: plural de “pião”

Palavra esperada: plural de "jornal"

Palavra esperada: plural de "troféu"

Palavra esperada: plural de "cartão"

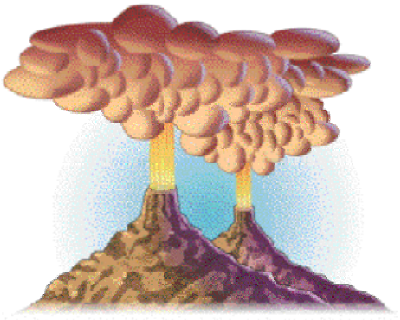
Palavra esperada: plural de "balão"

Palavra esperada: plural de "sinal"

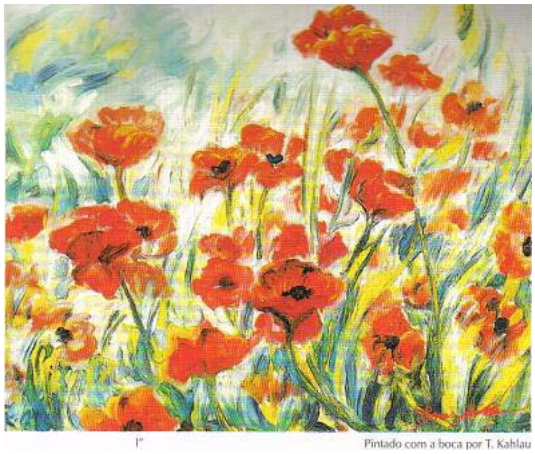
Palavra esperada: plural de “escorpião”

Palavra esperada: plural de “limão”

Palavra esperada: plural de “grau”



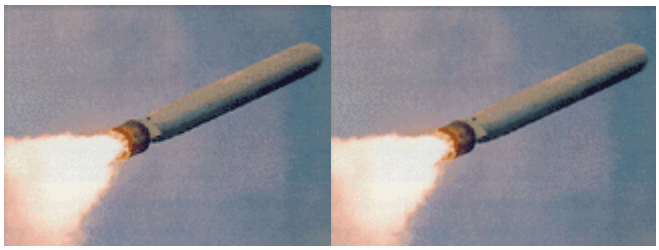
Palavra esperada: plural de “vulcão”



Palavra esperada: plural de “flor”

Palavra esperada: plural de “relógio”

Palavra esperada: plural de “degrau”

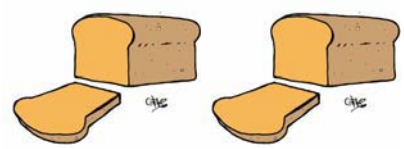


Palavra esperada: plural de “míssil”

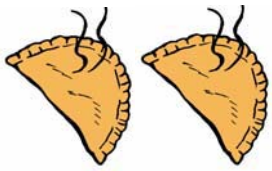
Palavra esperada: plural de “pneu”

Palavra esperada: plural de “quadro”

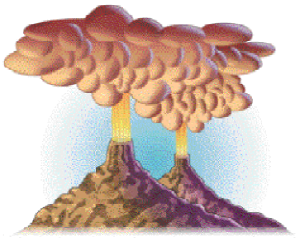
2. Leitura de Frases

1. Compramos alguns  integrais.

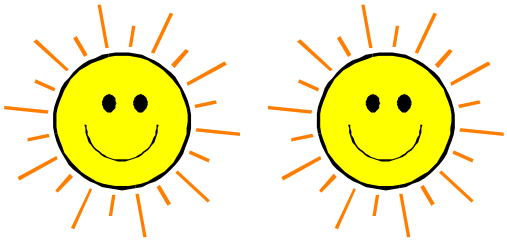
Palavra esperada: plural de "pão"

2. Meus  favoritos são de banana com canela.

Palavra esperada: plural de "pastel"

3. Os  da Califórnia sempre entram em erupção.

Palavra esperada: plural de "vulcão"

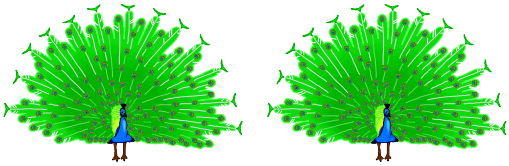
4. Será que existem dois  na nossa galáxia?

Palavra esperada: plural de "sol"

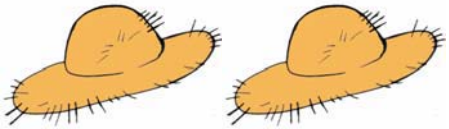
5. A medicina chinesa acredita que nossos

têm muitos pontos energéticos.

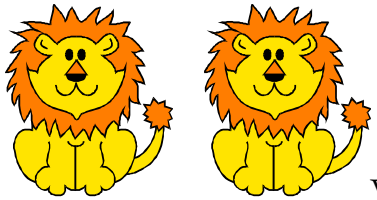
Palavra esperada: plural de "pé"

6. Chegaram dois  coloridos no zoológico.

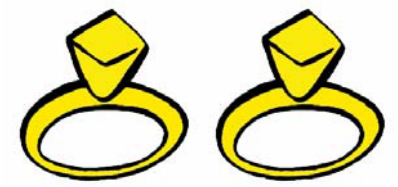
Palavra esperada: plural de "pavão"

7. Ainda estamos precisando de dois  de palha emprestados para a quadrilha.

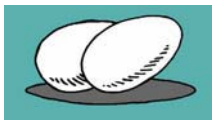
Palavra esperada: plural de "chapéu"

8. Os  vivem na África.

Palavra esperada: plural de "leão"

9. Quantos  você tem?

Palavra esperada: plural de "anel"

10. Muita gente gosta de comer  cozidos.

Palavra esperada: plural de "ovo"

11. Nosso time ganhou vários no torneio do ano passado.

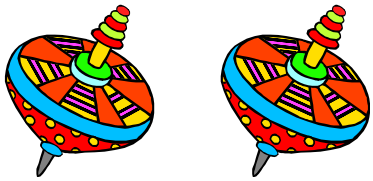
Palavra esperada: plural de "troféu"

12. As dele estavam muito ressecadas pelo frio.

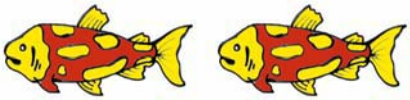
Palavra esperada: plural de "mão"

13. O pessoal da capoeira estava tocando os .

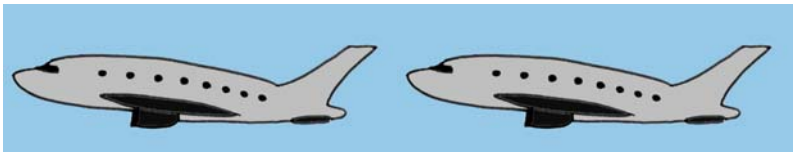
Palavra esperada: plural de "berimbau"

14. Crianças geralmente gostam de rodar  .

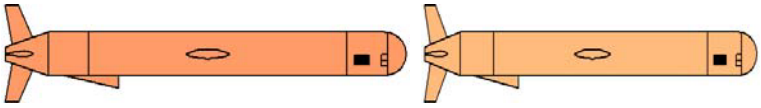
Palavra esperada: plural de "pião"

15. Os  de água doce são muito bons.

Palavra esperada: plural de "peixe"

16. Os  estavam voando muito alto.

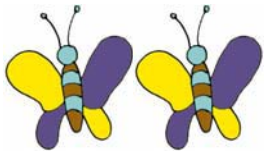
Palavra esperada: plural de "avião"

17. Os Estados Unidos lançaram vários  na Guerra contra o Iraque.

Palavra esperada: plural de "míssil"

18. Todos os de trânsito estavam apagados por causa da chuva.

Palavra esperada: plural de "sinal"

19. As  gostam de dias quentes e de céu azul.

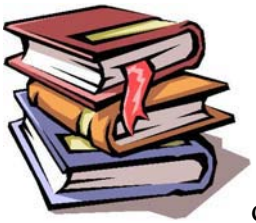
Palavra esperada: plural de "borboleta"

20. O carro derrapou porque os estavam muito carecas.

Palavra esperada: plural de "pneu"

21. À noite, os navios se guiam pelos .

Palavra esperada: plural de "farol"

22. Os  de literatura brasileira são interessantes.

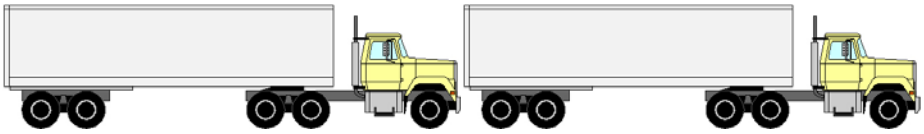
Palavra esperada: plural de "livro"

23. A maioria das pessoas não gosta de ler .

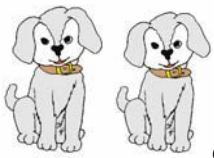
Palavra esperada: plural de "jornal"

24. Ontem à noite estava frio e o termômetro estava marcando 20 .

Palavra esperada: plural de "grau"

25.Dirigir  é muito arriscado nas estradas brasileiras.

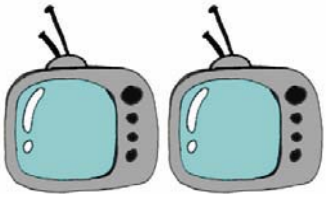
Palavra esperada: plural de "caminhão"

26. Ela tinha dois  enormes e muito bravos.

Palavra esperada: plural de "cachorro"

27. Resolvi reformar minha casa e comprei vários novos.

Palavra esperada: plural de "móvel"

28. As  são bem mais baratas hoje em dia.

Palavra esperada: plural de "televisão"

29. Esta escada tem muitos .

Palavra esperada: plural de "degrau"

30. Pedi a ele pra comprar dois novos, um pra mim e outro pra você.

Palavra esperada: plural de "funil"

3. Teste de Reação

1. Girafa
2. Menino
3. Alimento
4. Mercado
5. Batata
6. Vizinho
7. Árvore
8. Barato
9. Avental
10. Pão
11. Chapéu
12. Difícil
13. Degrau
14. Guardião
15. Judeu
16. Espanhol
17. Órgão
18. Mausoléu
19. Avião
20. Jirau
21. Acessível
22. Irmão
23. Gol
24. Ateu
25. Gentil
26. Anão
27. Humano
28. Seu
29. Bênção
30. União
31. Azul
32. Escrivão
33. Útil
34. Mau
35. Exposição
36. Anel
37. Domingo
38. Véu
39. Cristão
40. Europeu
41. Anzol
42. Mel
43. Opinião
44. Grau
45. Agradável
46. Mão
47. Museu
48. Sol
49. Vulcão
50. Meu
51. Religião
52. Lençol
53. Réu
54. Antigo
55. Pau
56. Alemão
57. Hospital
58. Responsável
59. Céu
60. Riqueza
61. Cão
62. Troféu
63. Sal
64. Limão
65. Infantil

Teste sobre Glidificação

1. Leitura de Frases[104]

Faz três dias que não pára de chover.

Sua irmã faz pós-graduação há quanto tempo?

O gás acabou.

Há uns dez anos atrás nós fomos a Foz do Iguaçu.

Ele tem um nariz tão grande!

Os urubus estavam voando no céu, bem em cima de casa.

Adorei a voz do seu namorado!

Chove demais no mês de janeiro.

A paz no Oriente Médio ainda vai demorar.

Ela parece ser uma pessoa muito feliz.

Comprei quatro maçãs, três cajus e dois abacaxis.

A luz na sua casa é 110 ou 220?

[104] Sublinhamos as palavras que, na "Leitura de Frases", seriam propensas à glidificação.

2. Plurais

1. Mesa
2. Véu
3. Árvore
4. Pé
5. Avião
6. Letra
7. Céu
8. Fotografia
9. Chuchu
10. Cipó
11. Cadeira
12. Planta
13. Borboleta
14. Janela
15. Acarajé
16. Flor
17. Rio
18. Caneta
19. Chá
20. Panela
21. Nu
22. Réu
23. Cama
24. Lâmpada
25. Sal
26. Chuva
27. Laranja
28. Gel
29. Pá
30. Folha
31. Moeda
32. Lua
33. Nó
34. Mel
35. Relógio
36. Sol
37. Livro
38. Céu
39. Café
40. Porta
41. Anel
42. Jiló

Anexo B – Tabelas referentes ao "Teste de Reação"

Neste Anexo, apresentamos as tabelas secundárias referentes ao "Teste de Reação". No Capítulo 6, "Análise de Dados", comentamos os resultados gerais quanto ao "Teste de Reação", mas não mencionamos o efeito das variáveis que já haviam sido selecionadas quando analisados os resultados gerais para cada grupo de plurais. Neste Anexo, faremos comentários mais detalhados e pontuais acerca de todos os fatores relevantes para o "Teste de Reação".

Primeiramente, analisaremos as tabelas referentes aos plurais em *–l*, começando pela variável **número de sílabas**:

Tabela 62: Efeito do número de sílabas no "Teste de Reação" para plurais em *–l*

Número de Sílabas	N	%	P.R.
Monossílabos	59/144	40,9	0.88
Polissílabos	39/502	7,7	0.11

Observamos, no "Teste de Reação", que os monossílabos em *–l* favoreceram as hesitações (0.88), enquanto os polissílabos as desfavoreceram (0.11). Na análise dos dados gerais para plurais em *–l* (Cf. Tabela 41), observamos que monossílabos em *–l* eram mais suscetíveis à adoção de plurais em ditongo em *–u* do que polissílabos. Talvez por isso tenha havido maior hesitação na classe de itens de uma só sílaba.

A seguir, analisamos os resultados referentes à **vogal precedente**[105] para os plurais em *–l*:

Tabela 63: Efeito da vogal precedente no "Teste de Reação" para plurais em *–l*

Vogal Precedente	N	%	P.R.
[a]	21/108	19,4	0.57
[ɛ]	**18/71**	**25,3**	**0.65**
[e]	6/108	5,5	0.25
[i]	18/144	12,5	0.44
[ɔ]	22/143	15,3	0.5
[o]	**9/36**	**25**	**0.65**
[u]	4/36	11,1	0.41

Nesta tabela, observamos que as vogais precedentes que mais causaram hesitações nos falantes foram [ɛ] e [o]. Com vogal precedente [ɛ] há o item "mel", um dos grandes líderes na adoção de plural em

[105] Para alcançar resultados de peso relativo para a vogal precedente nos plurais em ditongo em *–u*, foi necessário excluir as demais variáveis internas, caso contrário os resultados para as vogais eram categóricos. O mesmo ocorreu, também, nas Tabelas 39 e 40 (Cf. Capítulo 6, "Análise de Dados").

ditongo em *–u*. De forma similar, com a vogal [o] há apenas a palavra "gol", que apresentou números categóricos de plural em *–u*. Sendo assim, as maiores hesitações ocorreram justamente nas vogais precedentes cujas palavras são mais suscetíveis à adoção de plurais analógicos. Ou seja, parece que esse efeito da vogal precedente decorre, na verdade, de itens léxicos específicos que favoreceram a dúvida nos falantes.

Abaixo apresentamos os resultados para **estrutura morfológica**[106] das palavras terminadas em *–l*.

Tabela 64: Efeito da estrutura morfológica no "Teste de Reação" para plurais em *–l*

Estrutura Morfológica	N	%	P.R.
Sufixo	18/251	7,1	0.39
Não-Sufixo	80/395	20,2	0.60

Quanto à estrutura morfológica, observamos que o "Teste de Reação" apresenta resultados semelhantes àqueles alcançados para os plurais em *–l* como um todo (Cf. Tabelas 37 e 38). Observamos que os falantes hesitaram mais (0.60) quando a palavra não tinha sufixo e hesitaram menos (0.39) quando a palavra tinha sufixo. Mais uma vez, percebemos que a estrutura morfológica é uma referência importante para que o item léxico mantenha sua classe etimológica ou migre para a classe analógica. Itens com sufixo causam menos dúvidas nos falantes porque a existência de sufixo é uma característica marcante na classe de *–l*, o que faz com que haja baixa variabilidade em itens de estrutura morfológica complexa. No caso das palavras sem sufixo, elas apresentam estrutura semelhante à dos itens em ditongo em *–u*, já que estes, em sua maioria, não apresentam sufixo; sendo assim, palavras terminadas em *–l*, quando não têm sufixo, causam maiores dúvidas nos falantes.

A seguir, mostramos os resultados para os plurais em ditongo em *–u*, começando pela variável **estrutura morfológica**[107]:

Tabela 65: Efeito da estrutura morfológica no "Teste de Reação" para plurais em ditongo em *–u*

Estrutura Morfológica	N	%	P.R.
Sufixo	16/71	22,5	0.62
Não-Sufixo	99/536	18,4	0.36

Os resultados do "Teste de Reação" para a estrutura morfológica de itens em ditongo em *–u* são diferentes daqueles encontrados na análise geral dos dados de plurais em ditongo em *–u* (Cf. Tabela 49), já que, no total dos dados, palavras que tinham sufixo foram menos propensas a migrações de plural (0.27) do que palavras que não tinham sufixo (0.72). Os itens com sufixo talvez tenham provocado mais hesitações

[106] Nos cálculos de peso relativo para estrutura morfológica, removemos a variável número de sílabas porque ela estava interferindo na gradualidade dos resultados.

[107] A fim de alcançar resultados de peso relativo compatíveis com os números percentuais, excluímos a variável freqüência de ocorrência, porque ela estava interagindo com estrutura morfológica, impedindo correlatos adequados de porcentagem e peso relativo.

porque, conforme vimos no Capítulo 2, "Grupos de Plurais", há mais sufixos na classe de *–l* do que na de ditongo em *–u*, o que pode ter provocado dúvidas quanto ao plural adequado para a palavra terminada em *–u* com sufixo. Sendo a existência de sufixo uma forte referência para a classe em *–l*, os falantes podem ter apresentado dúvidas quando se depararam com itens em ditongo em *–u* que apresentavam sufixo.

Abaixo apresentamos o efeito da **vogal precedente** para as palavras terminadas em ditongo em *–u*:

Tabela 66: Efeito da vogal precedente no "Teste de Reação" para plurais em ditongo em *–u*

Vogal Precedente	N	%	P.R.
[a]	39/179	21,7	0.59
[ɛ]	50/214	23,3	0.58
[e]	26/214	12,1	0.32

Nesta tabela, observamos que os itens em ditongo em *–u* com vogal precedente [a] e [ɛ] favorecem ligeiramente as hesitações (0.59 e 0.58, respectivamente), enquanto os itens com vogal precedente [e] desfavorecem as hesitações (0.32). Este resultado é um pouco diferente daquele observado para os dados como um todo, porque, naquele caso, a vogal precedente [ɛ] favorecia bastante a adoção de plural em *–l* (0.83), a vogal [a] também favorecia (0.71) e a vogal [e] desfavorecia (0.07) (Cf. Tabela 48). No "Teste de Reação", é possível que as palavras "berimbau" e "jirau", que favoreceram plural em *–l*, tenham elevado as probabilidades de itens terminados com a vogal [a] sofrerem hesitações. Portanto, possivelmente há itens léxicos específicos favorecendo dúvidas nos seus plurais. O efeito da vogal precedente parece ser conseqüência dessas palavras individualmente.

Finalmente, na tabela a seguir, analisamos a interferência da **escolaridade** nos plurais em ditongo em *–u*:

Tabela 67: Efeito da escolaridade no "Teste de Reação" para plurais em ditongo em *–u*

Nível de escolaridade	N	%	P.R.
Fundamental	67/305	21,9	0.55
Superior	48/302	15,8	0.44

No "Teste de Reação", observamos que os falantes de Nível Fundamental também lideram as hesitações no grupo de ditongos em *–u* (peso relativo de 0.55 contra 0.44 dos falantes de nível superior). Percebemos, portanto, que os informantes com nível universitário têm mais certeza quanto aos plurais das palavras e, por isso, hesitam menos no momento de responder. Quando analisamos os dados gerais (Cf. Tabela 53), também observamos que os falantes de Nível Fundamental são os líderes nas migrações de plurais, enquanto os falantes de Nível Superior tendem a evitar tal fenômeno. Mais uma vez, percebemos que o fato de o indivíduo ter freqüentado mais a escola pode ter proporcionado um contato maior com a língua

escrita, modalidade em que as formas generalizadas de plural ocorrem menos. Conseqüentemente, falantes mais escolarizados apresentam menos hesitações quando são questionados sobre o plural de itens léxicos específicos.

Finalizamos aqui os comentários sobre as tabelas que não foram mencionadas durante a análise do "Teste de Reação", no Capítulo 6, "Análise de Dados".

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Administração
Baixar livros de Agronomia
Baixar livros de Arquitetura
Baixar livros de Artes
Baixar livros de Astronomia
Baixar livros de Biologia Geral
Baixar livros de Ciência da Computação
Baixar livros de Ciência da Informação
Baixar livros de Ciência Política
Baixar livros de Ciências da Saúde
Baixar livros de Comunicação
Baixar livros do Conselho Nacional de Educação - CNE
Baixar livros de Defesa civil
Baixar livros de Direito
Baixar livros de Direitos humanos
Baixar livros de Economia
Baixar livros de Economia Doméstica
Baixar livros de Educação
Baixar livros de Educação - Trânsito
Baixar livros de Educação Física
Baixar livros de Engenharia Aeroespacial
Baixar livros de Farmácia
Baixar livros de Filosofia
Baixar livros de Física
Baixar livros de Geociências
Baixar livros de Geografia
Baixar livros de História
Baixar livros de Línguas

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo