Pablo Arantes

Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala

Doutorado em Linguística

Universidade Estadual de Campinas
Instituto de Estudos da Linguagem

Ficha catalográfica elaborada pela Biblioteca do IEL – Unicamp

Arantes, Pablo.
Ar14i Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala / Pablo Arantes. – Campinas, SP: [s.n.], 2010.

Orientador: Plinio Almeida Barbosa.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Língua portuguesa - Ritmo. 2. Língua portuguesa - Brasil. 3. Língua portuguesa - Acentos e acentuação. 4. Percepção da fala. 5. Fonética. I. Barbosa, Plinio Almeida. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

oe/iel

Título em inglês: Integrating production and perception of secondary prominences in a dynamical perspective of rhythm

Palavras-chave em inglês (Keywords): Brazilian Portuguese – Rhythm; Brazilian Portuguese – Brazil; Portuguese language - Accents and accentuation; Speech perception; Phonetics.

Área de concentração: Linguística

Titulação: Doutor em Linguística

Banca Examinadora: Prof. Dr. Plinio Almeida Barbosa (orientador), Prof. Dr. Plinio Almeida Barbosa, Prof. Dr. João Antonio de Moraes, Profa. Dra. Maria Bernadete Marques Abaurre, Profa. Dra. Sandra Madureira Fontes e Prof. Dr. Edson Françozo.

Data da defesa: 04/02/2010

Programa de Pós-Graduação: Programa de Pós-Graduação em Linguística.

Pablo Arantes

Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala

Tese de Doutorado apresentada ao Instituto de Estudos da Linguagem como parte dos requisitos para obtenção do título de Doutor em Linguística.
Orientador: Plinio A. Barbosa

BANCA EXAMINADORA:

Plínio Almeida Barbosa

João Antonio de Moraes

Maria Bernadete Marques Abaurre

Sandra Madureira Fontes

Edson Françozo

Universidade Estadual de Campinas
Instituto de Estudos da Linguagem
Campinas, fevereiro de 2010

Aos meus pais, Marilene e Claudio

Agradecimentos

A todos aqueles que, com os impostos gerados pelo seu trabalho, tornaram possível a existência das excelentes escolas públicas em que sempre estudei. E também a todos os que lutam para que elas continuem públicas e gratuitas.

Ao meu orientador, Plinio Barbosa, pela orientação rigorosa e dedicada. Pelo encorajamento constante e entusiasmo com que sempre discutiu nosso trabalho conjunto. Por dividir seu conhecimento de forma generosa. Por ter aguçado minha capacidade analítica como linguista e ter contribuído para o meu domínio da estatística e da programação, que se tornaram duas fontes inesperadas de prazer intelectual para mim. E por fazer tudo isso com paciência, humildade, franqueza, ética e espírito agregador. Que possamos ser parceiros por muitos anos mais.

Ao professor Edson Françozo, cujo estilo socrático em aula e no laboratório incutiu em mim o gosto pela prática da ciência empírica e pela epistemologia. Pelas aulas de lógica no segundo semestre da graduação que marcaram a turma pelo resto do curso. Pelas férias de verão passadas no laboratório tentando fazer a rede neural Palau falar. Pelas sempre esperadas pausas no trabalho para as conversas acompanhadas de café. Pela amizade fora do ambiente acadêmico e ajuda em momentos pessoais graves.

Aos professores Rodolfo Ilari, Wilmar D'Angelis e Ana Luiza Navas por aulas sempre inspiradas.

Aos membros da banca por terem aceito o convite.

À Aglael Gama Rossi, Bernadete Abaurre, e João Antônio de Moraes, membros das bancas de qualificação, cuja leitura ajudou a aprimorar o resultado final da tese.

Aos funcionários da Unicamp que fazem o dia-a-dia do trabalho acadêmico possível.

À Fundação de Amparo à Pesquisa do Estado de São Paulo pela bolsa de mestrado e posteriormente doutorado direto que permitiram que eu me dedicasse inteiramente à pesquisa.

Aos colegas do Grupo de Estudos de Prosódia da Fala, Ana Carolina Constantini, Ana Cristina Fricke Matte, Alexsandro Meireles, Aveliny Lima, Leandro Silveira, Luciana Lucente e Sandra Merlo pela camaradagem.

Ao movimento do *software* livre e de código aberto, que torna possível a existência das ferramentas computacionais livres de alta qualidade que usei para a produção desta tese. Em particular a Paul Boersma e David Weenink pela criação do Praat e a R Foundation for Sta-

tistical Computing pela criação do ambiente de computação estatística R.

Aos locutores dos experimentos de produção, Antonio Pessotti e Celio Costa, e aos muitos sujeitos dos experimentos de percepção pelo tempo e paciência.

Aos amigos de Campinas cuja companhia tornou a vida mais divertida e significativa: Alan Carneiro, Fernando Ferrone, Giovani Ronchi, Gustavo Conde, Leonardo Oliveira, Lígia Gomes, Lou-Ann Kleppa, Orlando Coelho, Ricardo Garcia, Rogério Ribeiro e Ruy Braz.

Ao Antonio Barros, Celio Costa, Laudino Rocés e Renato Basso, minha segunda família em Campinas.

A meus pais, Marilene e Claudio por terem dedicado tanto de suas vidas na formação dos filhos, pelo incondicional apoio emocional e confiança. Às minhas irmãs Carolina e Julia, minha avó Ziza e tia Marisa pelo carinho e encorajamento constantes. Tudo o que tenho de bom vem de vocês.

À Maria Luiza pela cumplicidade e companheirismo em nossa vida compartilhada. E acima de tudo porque "o amor é como uma corruíra no jardim — de repente ela canta e muda toda a paisagem".

Nossa lavagem de louça é como nossa linguagem, disse Niels [Bohr]. Temos água suja e panos de prato sujos e, no entanto, conseguimos deixar limpos os pratos e os copos. Também na linguagem, temos de trabalhar com conceitos pouco claros e com um tipo de lógica cujo alcance é restrito e desconhecido. No entanto, nós a usamos para introduzir clareza em nossa compreensão da natureza.

Werner Heisenberg, *A parte e o todo*

Resumo

Nesta tese são investigados dois temas: o acento secundário em português brasileiro e a integração entre a produção e a percepção do ritmo. Os dois temas são desenvolvidos tendo como pano de fundo o modelo dinâmico de produção do ritmo da fala de Barbosa (2006 e 2007). O fenômeno descrito como acento secundário é associado à ocorrência de proeminências na cadeia de sílabas pré-tônicas em palavras polissilábicas. O interesse pelo assunto decorre das disputas existentes na literatura sobre a natureza (e mesmo existência) destes acentos e os princípios que determinam sua distribuição e, ainda, da limitada quantidade disponível de dados experimentais para resolver essas disputas. Foram realizados dois experimentos de produção com o objetivo de descrever, em palavras polissilábicas com mais de duas pré-tônicas, os padrões apresentados por diversos parâmetros acústicos que são tradicionalmente associados à expressão do acento. Nos corpora dos experimentos foram manipuladas, entre outras variáveis, o tamanho da cadeia de pré-tônicas, a posição sintática e o *status* referencial das palavras-chave. Os parâmetros acústicos investigados foram duração, frequência fundamental, qualidade vocálica e ênfase espectral. Os resultados sugerem que há evidência para a manifestação de uma proeminência nas posições iniciais da cadeia de pré-tônicas. Essa proeminência, que tem um caráter opcional e dependente de falante, se manifesta acusticamente por meio do alongamento relativo gradiente das unidades de vogal a vogal em início de enunciado e de picos na curva de frequência fundamental alinhados às primeiras sílabas pré-tônicas da cadeia. O número de sílabas pré-tônicas é, dentre os fatores controlados nos corpora, o que mais tem influência sobre a manifestação da proeminência inicial. Simulações com o modelo dinâmico mostram que ele pode reproduzir os padrões de alongamento inicial observados nos dados coletados, o que pode ser tomado como evidência para a adequação de sua formulação corrente. Uma proposta de tratamento comum da produção e da percepção do ritmo em um modelo unificado que estende a noção de acoplamento entre osciladores para a relação entre a produção e a percepção foi investigada por meio de dois experimentos de percepção. Procurou-se obter por meio dos experimentos evidência comportamental para a proposição, implícita na proposta, segundo a qual a percepção da estruturação temporal pelo ouvinte é afetada pela maneira como a temporalidade foi produzida pelo falante. Nos experimentos, cliques foram inseridos junto aos *onsets* vocálicos ao longo de dois grupos acentuais consecutivos e a tarefa dada aos sujeitos era detectar os cliques. Os dados da latência de detecção dos cliques mostram que, ao longo do grupo acentual, os tempos de reação diminuem, tendendo a aumentar nas imediações da fronteira do grupo acentual, marcada por um pico no contorno de duração da frase. Análises de regressão indicam que as mudanças de duração e frequência fundamental ao longo do grupo acentual explicam em torno de 50% da variabilidade dos tempos de reação. Os resultados constituem evidência para a proposta de integração delineada dado que a relação encontrada nos dados pode ser interpretada como o resultado de acoplamento entre temporalidade produzida e percebida.

Palavras-chave: Português brasileiro, acento secundário, ritmo, percepção da fala.

Abstract

The thesis investigates two main subjects: secondary stress in Brazilian Portuguese (BP) and the link between rhythm production and rhythm perception. Both subject matters are developed in light of Barbosa's (2006, 2007) coupled-oscillator model of rhythm production. Interest in secondary stress in BP is fostered by the fact that its existence, nature and distribution patterns are disputed in the literature. Also, the few existing experimental data do not help to settle the issues. Two production experiments are reported that aimed at describing the patterns of acoustical parameters traditionally considered correlates of stress observed in target words with at least two prestressed syllables. The experiments' corpora manipulated target word variables such as the number of prestressed syllables, syntactic position and referential status. Duration, fundamental frequency, vowel quality and spectral emphasis were measured. Results suggest that there's evidence of optional and speaker dependent initial prominence marking. Lengthening of utterance-initial syllable-sized units and fundamental frequency peaks also aligned to initial syllables are the main correlates of initial prominence. The number of target words' prestressed syllables has the strongest influence in initial prominence marking among the variables manipulated in the corpora. Simulations run with the coupled-oscillator model indicate that it is able to generate patterns of initial lengthening similar to the ones observed in the experiment. We outlined a proposal for a unified approach to speech rhythm that extends the notion of coupling and entrainment to the speaker-hearer relationship. Two perception experiments were run to test the implicit proposition that rhythm perception is the result of listeners getting entrained by speaker's produced timing. In the experiment, clicks were inserted near each vowel onset along two consecutive stress groups in test sentences. Each test sentence had one click and there were as many test sentence copies as vowel onset in the two stress groups. Subjects were instructed to press a button as soon as they heard the click upon listening to test sentences. Results show that the closer the click is to the stress group boundary (signaled in the speech signal by a peak in normalized duration) the faster the RT is. The crucial result concerning the speaker-hearer entrainment hypothesis, though, is that RT slows down after the stress group boundary, resuming its decrease afterwards, suggesting that the listener is sensitive to the effects of the boundary on the temporal structure of the signal being heard. Regression analysis points out that previous intervocalic interval duration (relative to the interval the click is associated to) and the derivative of the median fundamental frequency over the intervocalic interval account for nearly 50% of the variance in detection latency times. Overall the results seem to provide initial evidence to the speaker-listener entrainment hypothesis by showing that important events in the speech signal like a boundary have an impact on listeners' flow of attention along the sentence. Also, results single out the importance of vowel onsets as a carrier of prosodic structure and its role as a common currency between production and perception.

Keywords: Brazilian Portuguese, secondary stress, rhythm, speech perception.

Sumário

Lista de Figuras **xix**

Lista de Tabelas **xxi**

Introdução geral **1**

1 O modelo dinâmico do ritmo **3**
1.1 Caracterização geral . . . 3
1.2 Formulação matemática do modelo . . . 5
1.3 Acento secundário no quadro dinâmico . . . 6
1.4 Integrando produção e percepção do ritmo . . . 8

2 O acento secundário **13**
2.1 Abordagens tradicionais . . . 14
2.2 Abordagem estruturalista: Mattoso Câmara Jr. . . . 15
2.3 Abordagens gerativistas . . . 15
2.4 Estudos experimentais em PB . . . 16
2.5 O fenômeno em outras línguas românicas . . . 18
2.5.1 Espanhol . . . 18
2.5.2 Italiano . . . 22
2.5.3 Francês . . . 25
2.6 Discussão . . . 28

3 Estudo de produção I **31**
3.1 Metodologia . . . 31
3.1.1 Variáveis independentes . . . 31
3.1.2 Frases-veículo . . . 33
3.1.3 Palavras-alvo . . . 34
3.1.4 Procedimentos de gravação . . . 35
3.1.5 Variáveis dependentes . . . 35
3.1.6 Análise estatística . . . 38
3.2 Resultados: Duração . . . 38
3.2.1 Influência dos fatores PS e d_σ . . . 39
3.2.2 Influência da natureza das vogais (VR) . . . 46
3.2.3 Influência da sílaba inicial ES . . . 49

3.2.4 Discussão parcial . . . 55
3.3 Resultados: Distribuição de acentos frasais . . . 56
3.3.1 Grupo A . . . 57
3.3.2 Grupo B . . . 57
3.3.3 Grupo C . . . 58
3.3.4 Grupo D . . . 58
3.3.5 Discussão parcial . . . 59
3.4 Resultados: Frequência fundamental . . . 59
3.4.1 Análise por valor no ponto médio . . . 59
3.4.2 Distribuição de pontos extremos H . . . 67
3.4.3 Discussão parcial . . . 69
3.5 Configuração formântica . . . 69
3.5.1 Vogal /a/ . . . 69
3.5.2 Vogal /i/ . . . 72
3.5.3 Discussão parcial . . . 74
3.6 Resultados: ênfase espectral . . . 75
3.6.1 Palavra *patarata* . . . 75
3.6.2 Palavra *jaratacaca* . . . 77
3.6.3 Discussão parcial . . . 81

4 Estudo de produção II **83**
4.1 Variáveis lexicais . . . 83
4.1.1 Obtenção da frequência de uso . . . 85
4.1.2 Definindo faixas de frequência . . . 85
4.1.3 Frequência de uso e ritmo . . . 86
4.2 Variáveis sintáticas . . . 87
4.3 Variável discursiva . . . 88
4.4 Metodologia . . . 89
4.4.1 Procedimentos de gravação . . . 89
4.4.2 Variáveis dependentes . . . 90
4.5 Resultados: duração . . . 90
4.5.1 Efeito da variável sintática . . . 91
4.5.2 Efeito da variável discursiva . . . 100
4.6 Resultados: frequência fundamental . . . 110
4.6.1 Tamanho da palavra-chave . . . 112
4.6.2 Variável sintática . . . 112
4.6.3 Efeito da variável discursiva . . . 116
4.7 Simulação com o modelo de ritmo . . . 121
4.7.1 Procedimento . . . 121
4.7.2 Grupos acentuais simulados . . . 122
4.7.3 Resultados . . . 122

5 Experimentos de percepção **125**
5.1 Experimento de percepção I . . . 128
5.1.1 Estímulos experimentais . . . 129
5.1.2 Variáveis . . . 130
5.1.3 Procedimentos . . . 131
5.1.4 Análise estatística . . . 131
5.1.5 Resultados . . . 132
5.1.6 Discussão . . . 134
5.2 Experimento de percepção II . . . 134
5.2.1 O efeito lombada . . . 136
5.2.2 Estímulos experimentais . . . 138
5.2.3 Procedimento experimental . . . 138
5.2.4 Análise estatística . . . 140
5.2.5 Resultados . . . 140
5.3 Discussão . . . 144

Conclusão **147**

A Algoritmos para análise da duração **151**
A.1 Normalização da duração por z-score . . . 151
A.1.1 Justificativa . . . 151
A.1.2 Implementação . . . 152
A.2 Suavização do contorno de duração . . . 152
A.3 Detecção dos picos de duração . . . 152
A.4 Exemplo de aplicação . . . 153

B Algoritmo para análise de f_0 **155**
B.1 Justificativa . . . 155
B.2 Implementação . . . 157

C Experimento de produção II **161**
C.1 Lista das frases-veículo do *corpus* . . . 161

Referências bibliográficas **165**

Lista de Figuras

1.1 Diagrama do modelo dinâmico do ritmo (Barbosa 2006) 4
1.2 Integração da percepção no modelo de produção do ritmo 9

3.1 Duração média de sílabas e unidades V-V – grupo A 41
3.2 Duração média de sílabas e unidades V-V – grupo B 42
3.3 Duração média de sílabas e unidades V-V – grupo C 44
3.4 Duração média de sílabas e unidades V-V – grupo D 45
3.5 Efeito da variável VR sobre as sílbas – grupo A 47
3.6 Efeito da variável VR sobre as unidades V-V – grupo A 48
3.7 Efeito da variável VR sobre as sílabas – grupo B 50
3.8 Efeito da variável VR sobre as unidades V-V – grupo B 51
3.9 Efeito da variáveil ES – grupo A 53
3.10 Efeito da variáveil ES – grupo B 54
3.11 Distribuição dos acentos frasais – grupo A 60
3.12 Distribuição dos acentos frasais – grupo B 61
3.13 Distribuição dos acentos frasais – grupo C 62
3.14 Distribuição dos acentos frasais – grupo D 63
3.15 F_0 no ponto médio – grupo A 64
3.16 F_0 no ponto médio – grupo B 65
3.17 F_0 no ponto médio – grupo C 65
3.18 F_0 no ponto médio – grupo D 66
3.19 Extremos H e R por posição — todos os grupos 68
3.20 Valores de F_1 de /a/ e correlação com duração — *patarata* 71
3.21 Valores de F_1 de /a/ e correlação com duração — *jaratacaca* 73
3.22 Valores médios de F_1 de /i/ — *dirigibilidade* 74
3.23 Ênfase espectral e correlação com duração — *patarata* 79
3.24 Ênfase espectral e correlação com duração — *jaratacaca* 80

4.1 Frequência de uso e número de sílabas da palavra (Zipf 1965) 85
4.2 Duração de V-V e sílabas – variável sintática – grupo P2 93
4.3 Duração V-V e sílabas – variável sintática – grupo O2 94
4.4 Duração V-V e sílabas – variável sintática – grupo P3 96
4.5 Duração V-V e sílabas – variável sintática – grupo O3 97
4.6 Duração V-V e sílabas – variável sintática – grupo P4 99

4.7 Duração V-V e sílabas – variável sintática – grupo O4 101
4.8 Duração V-V e sílabas – variável discursiva – grupo P2 103
4.9 Duração V-V e sílabas – variável discursiva – grupo O2 104
4.10 Duração V-V e sílabas – variável discursiva – grupo P3 105
4.11 Duração V-V e sílabas – variável discurisva – grupo O3 107
4.12 Duração V-V e sílabas – variável discurisva – grupo P4 108
4.13 Duração V-V e sílabas – variável discurisva – grupo O4 111
4.14 Pontos extremos H e R: posição sintática inicial 113
4.15 Pontos extremos H e R: posição sintática medial 115
4.16 Pontos extremos H e R: referentes novos 118
4.17 Pontos extremos H e R: referentes velhos 120
4.18 Contornos duracionais simulados 123

5.1 Gráfico quantil–quantil da amostra de TR— experimento de percepção I 133
5.2 Médias de TR por unidade V-V — experimento de percepção I 135
5.3 Efeito lombada em Martin (1986) 137
5.4 Caracterização acústica de um dos estímulos do experimento II 139
5.5 Tempo de reação – grupo Pr2 – experimento de percepção II 141
5.6 Tempo de reação – grupo Pr3 – experimento de percepção II 142
5.7 Dispersão do TR em função da duração e f_0 — experimento II 143

A.1 Exemplo da aplicação dos algoritmos de análise da duração 154

B.1 Estilização da f_0 por pontos 156
B.2 Pontos extremos em um contorno de f_0 159
B.3 Alinhamento entre unidades V-V e pontos extremos 160

Lista de Tabelas

2.1 Esquema dos estímulos em Hoskins (1994). 27

3.1 Palavras-chave do experimento de produção I 34
3.2 Grupos de palavras-chave 39
3.3 Agrupamentos prosódicos – grupo A 57
3.4 Agrupamentos prosódicos – grupo B 58
3.5 Agrupamentos prosódicos – grupo C 58
3.6 Agrupamentos prosódicos – grupo D 59
3.7 Pontos extremos H 67
3.8 *Patarata*: correlações entre duração e F_1 70
3.9 *Jaratacaca*: correlações entre duração e F_1 72
3.10 *Patarata*: correlações entre duração e ênfase espectral (*B2–B1*) 76
3.11 *Patarata*: correlações entre duração e ênfase espectral (*B4–B3*) 77
3.12 *Jaratacaca*: correlações entre duração e ênfase espectral (*B2–B1*) 78
3.13 *Jaratacaca*: correlações entre duração e ênfase espectral (*B4–B3*) 78

4.1 Palavras-chave do experimento de produção II 84
4.2 Grupos de palavras-chave do experimento de produção II 91
4.3 Duração da palavra-chave e variável pragmático-discursiva 110
4.4 Efeitos globais do tamanho da palavra sobre f_0 112
4.5 Pontos extremos H: posição inicial 114
4.6 Pontos extremos H: posição medial 114
4.7 Pontos extremos L: posição final 114
4.8 Efeitos globais da variável discursiva sobre f_0 117
4.9 Pontos extremos H: referentes novos 117
4.10 Pontos extremos H: referente velho 119

Introdução

O assunto de que trata esta tese é o ritmo da fala e ele será desenvolvido a partir da investigação de dois temas principais: o acento secundário em português brasileiro (doravante PB) e a relação entre a produção e a percepção do ritmo da fala. Estes temas podem, em uma primeira aproximação, não parecer estreitamente relacionados, mas a ligação entre eles se torna aparente quando um terceiro elemento entra em cena. O elemento que promove a ligação entre os dois e que funciona como um pano de fundo contra o qual contrastar os dois temas particulares é o modelo dinâmico do ritmo desenvolvido por Barbosa (2006 e 2007) e, de forma mais geral, o quadro teórico e metodológico em que o modelo se enquadra. Os dois temas centrais da tese são, por assim dizer, recrutados em função de sua importância para o desenvolvimento do modelo, muito embora sejam questões linguísticas gerais o suficiente para terem importância independentemente de sua relação com o empreendimento do modelo.

O estudo do acento secundário tornou-se um dos temas da tese por ser um campo de provas para o modelo dinâmico do ritmo. Uma parte importante da literatura sobre o assunto afirma que os acentos secundários manifestam-se de forma binária ao longo da cadeia das sílabas pré-tônicas. O modelo dinâmico trata da produção do ritmo, gerando contornos duracionais simulados e, por força de sua atual formulação, os contornos produzidos por ele não apresentam padrões que poderiam ser chamados de binários. Essa característica do modelo, a previsão de não ocorrência de proeminências duracionais binárias, conflita, portanto, com a existência, assumida por certas análises, de acentos secundários atribuídos de forma binária. Uma vez que os dados apresentados pela maior parte das análises do fenômeno são baseadas na intuição linguística e este tipo de dado é incomensurável com as previsões empíricas produzidas pelo modelo, impôs-se então um trabalho considerável de descrição fonética para que se tornasse possível avaliar em bases empíricas o conflito entre a análise binarista a respeito do acento secundário e as previsões feitas pelo modelo. Essa é uma questão muito importante pois seria preciso revisar a formulação atual do modelo caso ele não desse conta de reproduzir os padrões apresentados pelos dados de produção.

O segundo tema, a relação entre a produção e a percepção do ritmo, encontrou lugar na tese em função das possibilidades de desdobramento do modelo dinâmico do ritmo. Como foi mencionado no parágrafo anterior, o modelo trata da produção do ritmo e seria interessante explorar o potencial do modelo para integrar as dimensões da produção e da percepção, uma vez que nenhuma das duas dimensões sozinhas parece esgotar o fenômeno do ritmo. O quadro mais geral em que o modelo de Barbosa (2006) se enquadra nuança as fronteiras entre saber e fazer, permitindo assim que se diluam também as fronteiras entre produzir e perceber. De fato, os mesmos mecanismos empregados por Barbosa (2006) em seu modelo de produção, a indução

(*entrainment*) e acoplamento entre osciladores, recursos do repertório teórico e metodológico da teoria dos sistemas dinâmicos, podem também ser empregados na tentativa de tratar a percepção, como pretendeu fazer McAuley (1995) com seu modelo de percepção de padrões temporais. O próprio modelo de referência inspira-se em McAuley (1995), e Barbosa (2006) sugere os traços gerais de uma abordagem que integra produção e percepção em um quadro comum. Neste panorama, a contribuição específica e mais original da tese foi a obtenção de evidência comportamental que possa corroborar os pressupostos assumidos pela proposta.

A organização da tese segue aproximadamente a ordem de apresentação do conteúdo assim como descrito até aqui na Introdução. O capítulo 1 apresenta o modelo dinâmico do ritmo, situando-o brevemente no contexto mais amplo das aplicações da teoria dos sitemas dinâmicos ao estudo da fala. Nesse mesmo capítulo trata-se da questão do acento secundário e da incompatibilidade da formulação atual do modelo com a hipótese de um acento secundário binário e das possibilidades que o modelo tem para lidar com um cenário em que acentos secundários binários devem ser implementados (seção 1.3). A seção 1.4 faz uma apresentação da proposta de abordagem integrada da produção e da percepção do ritmo da fala desenvolvida a partir das sugestões de Barbosa (2006) e indica o tipo de evidência que poderia corroborar a proposta.

O capítulo 2 lida com fenômeno do acento secundário, expõe diversas posições a respeito da natureza e da manifestação do acento secundário encontradas na literatura sobretudo na que diz respeito às línguas românicas, além de resenhar trabalhos de descrições acústicas do fenômeno. Os capítulos 3 e 4 apresentam os dois experimentos de produção elaborados para fornecer uma descrição fonética do acento secundário o mais rica possível em termos de parâmetros acústicos investigados. Além de contribuir para a descrição linguística de um fenômeno prosódico do PB pouco documentado do ponto de vista acústico, os dados são comparados às previsões feitas pelo modelo dinâmico do ritmo.

No capítulo 5 são descritos dois experimentos de percepção elaborados com a finalidade de prover evidência favorável para a hipótese segundo a qual a percepção do ritmo da fala é mediada pela maneira estruturada como se dá a sucessão dos intervalos entre *onsets* vocálicos ao longo da frase. Essa evidência é importante, pois aponta como esses intervalos são usados como uma unidade rítmica comum entre a produção e a percepção.

Capítulo 1

O modelo dinâmico do ritmo

1.1 Caracterização geral

O modelo de ritmo que será tomado como referência (para a versão mais atual ver Barbosa 2006 e 2007; para o desenvolvimento ver Barbosa 1996, 2001 e 2002a e Barbosa e Madureira 1999) faz parte da tentativa de construção de uma perspectiva dinâmica da produção da fala que, juntamente com outros autores (Browman e Goldstein 1989 e 1990 e Saltzman e Munhall 1989), busca na teoria dos sistemas dinâmicos (Kelso 1995) uma resposta para a questão da comensurabilidade entre o físico e o abstrato na produção da fala.

A manobra metodológica fundamental nesta perspectiva é a tentativa de redução formal do "saber" linguístico (associado tradicionalmente às representações fonológicas fundadas em primitivos estáticos e discretos, como o fonema ou o traço distintivo) e do "fazer" (associado à mecânica articulatória do trato vocal envolvida na implementação fonética) aos mesmos princípios passíveis de formalização pela teoria matemática dos sistemas dinâmicos (Kelso, Saltzman e Tuller 1986). No que concerne a produção da fala, este intento é conseguido quando se considera o gesto acústico-articulatório a unidade fônica autônoma mínima e formalizando-o por meio da equação dinâmica que descreve, em função do tempo, o comportamento de um sistema massa–mola com amortecimento crítico, que matematicamente equivale à trajetória assintótica de variáveis do trato vocálico (Browman e Goldstein 1986). Assim definido, o gesto é dotado de uma estrutura espacio-temporal, podendo tanto variar em magnitude relativa quanto em grau de sobreposição relativamente a outros gestos, explicando fenômenos comuns na gramática das línguas que, em perspectivas que tomam primitivos estáticos, são tratados como apagamentos ou inserções[1].

Já em relação ao ritmo da fala, a tarefa do quadro de referência dinâmico é explicitar os princípios que, atuando em conjunto, concorrem para sua organização temporal. Em Albano (2001), considera-se um nível intrínseco, responsável pelo controle das relações de fase entre os gestos que formam um item do léxico (representadas no que se chama *pauta gestual* a partir da especificação dos parâmetros das variáveis do trato). Ao mesmo tempo, considera-se um nível de organização atrelado a domínios superiores que modifica a temporalidade intrínseca,

[1] Cf. Albano (2001:68–96) para o caso da africação do [t] antes de [i], "iotização" de vogal antes de [s] de coda e inserção de [i] em encontros consonantais não envolvendo líquidas.

atuando como um marcador da temporalidade extrínseca.

Barbosa e Madureira (1999) apresentam uma proposta, incorporada por Albano (2001), para tratar a relação entre estas duas ordens de temporalidade no quadro da teoria dos sistemas dinâmicos na forma de uma hierarquia de osciladores acoplados (Strogatz e Stewart 1994, Kelso 1995 e McAuley 1995), através das noções de *indução*[2] de período (processo por meio do qual um oscilador muda progressivamente o seu período para que ele se aproxime do período de um outro oscilador por meio da influência deste) e de fase (sincronização, normalmente, do início do ciclo de um oscilador com o início do ciclo de um outro por meio da ação de um deles, que age como estímulo externo). A versão mais recente deste modelo, apresentado em Barbosa (2006 e 2007) pode ser vista na figura 1.1.

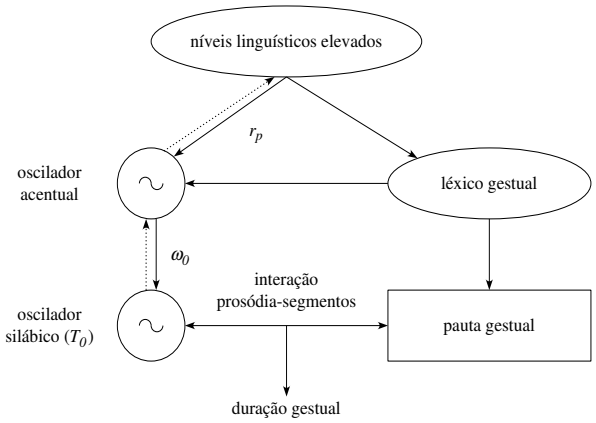


Figura 1.1: Diagrama do modelo dinâmico do ritmo na formulação apresentada em Barbosa (2006 e 2007).

Na versão do modelo apresentada em Barbosa (2006), propõe-se a existência de dois osciladores linguísticos abstratos. O primeiro deles é o oscilador a ser induzido, chamado de *oscilador silábico*, cuja ação é produzir a primeira especificação de duração da sequência de unidades do tamanho da sílaba que formam um dado enunciado. O outro é chamado de *oscilador acentual* e especifica a sequência, inicialmente eurrítmica, de acentos frasais em um enunciado. A ação indutora do oscilador acentual tem como resultado o aumento progressivo do período do oscilador silábico e atinge seu grau máximo quando coincidem os inícios de ciclo dos osciladores acentual e silábico. A batida do oscilador silábico já induzido interage com a pauta gestual, especificando o início (*onset*) dos gestos vocálicos. O modelo, assim como proposto, trata também da relação dos tempos intrínseco e extrínseco, uma vez que a coordenação entre os gestos vocálicos e consonantais é considerada uma propriedade da pauta gestual lexical, enquanto a organização dos gestos do tamanho da sílaba, cujo ponto de ancoragem é o

[2]Os termos *entrainment* e *entrained*, encontrados na literatura a respeito de osciladores acoplados, serão aqui traduzidos por *indução* e *induzido*.

onset da vogal, fica a cargo do sistema de osciladores acoplados, atuando de maneira extrínseca. Nesta perspectiva, a sílaba clássica só se faria especificada no léxico, na forma de relações de fase e magnitude relativa entre gestos tão estáveis que se poderia dizer que estão "congeladas". O modelo dá conta, assim, da emergência das propriedades de silabicidade e acentuação, esta última culminando em um acento lexical marcado frasalmente.

1.2 Formulação matemática do modelo

O acoplamento de período é implementado matematicamente por uma função de sincronismo $s(n)$, definida para cada unidade V-V de um grupo acentual arbitrário. Antes de apresentar a função, apresentaremos algumas convenções. Primeiramente, imaginemos um grupo acentual, "A macacada", composto de quatro unidades V-V[3]: |**a m**|**a c**|**a c**|**a d**|a). O símbolo N será usado para indicar o número total de unidades V-V no grupo acentual.

$$\underbrace{\overbrace{\mathrm{V}-\mathrm{V}_0}^{|\mathrm{am}|}\ \overbrace{\mathrm{V}-\mathrm{V}_1}^{|\mathrm{ac}|}\ \overbrace{\mathrm{V}-\mathrm{V}_2}^{|\mathrm{ac}|}\ \overbrace{\mathrm{V}-\mathrm{V}_3}^{|\mathrm{ad}|}}_{N=4}$$

O símbolo n ficará reservado para indicar o índice de uma unidade V-V qualquer tomada entre as N presentes no grupo acentual de interesse. Assim, os valores $n = 0$ e $n = N - 1$ denotam, respectivamente, os índices da primeira e da última unidades V-V num grupo acentual qualquer.

A equação 1.1 abaixo especifica a forma que a função $s(\cdot)$ adquire em três situações: $s(0)$, primeiro ciclo V-V do grupo acentual; $s(N-1)$, último ciclo de cada grupo acentual e $s(n)$, demais valores de n.

$$s(n) = \begin{cases} w_0 \cdot \exp(-N+2) & \text{se } n = 0, \\ \exp(-5,81 + 0,016 \cdot T_0) & \text{se } n = N-1, \\ (1-w_0) \cdot s(n-1) + w_0 \cdot \exp(-N+n+2) & \text{se } 0 < n < N-1. \end{cases} \tag{1.1}$$

Neste conjunto de equações, w_0 representa a força de acoplamento relativa entre os osciladores e T_0 é o período do oscilador silábico desacoplado, ou seja, sem a influência do oscilador acentual. A forma exponencial das funções foi determinada empiricamente, através de correlação não-linear, conforme procedimento relatado em Barbosa (2002a).

A mudança no valor do período do oscilador silábico para uma unidade V-V de índice n é calculada a partir da equação de diferenças finitas apresentada em 1.2.

$$\Delta T(n) = \alpha \cdot T(n-1) \cdot s(n) \cdot i(m) - \beta \cdot [T(n-1) - T_0] \cdot i(m-1) \tag{1.2}$$

Na equação 1.2, α representa a taxa de indução, que influencia a taxa de elocução, T é o valor do período do oscilador no instante considerado, $s(n)$ é a função de sincronismo mostrada na equação 1.1, $i(m)$ é a magnitude (especificada por um número real positivo) do próximo pulso do oscilador acentual, que é o elemento indutor do acoplamento para os ciclos do oscilador silábico

[3]Para simplificação, utilizaremos aqui uma representação ortográfica.

que o antecem. Como descrito anteriormente, o período do oscilador silábico induzido sofre a influência crescente do oscilador acentual, atingindo um valor culminante nas proximidades da batida do mesmo. Após a batida, ocorre o *resetting* do período, isto é, a duração do oscilador silábico tende a voltar ao que era antes da influência indutora do oscilador acentual, isto é, T_0, o valor do período de repouso. O segundo termo da equação 1.2 implementa este *resetting* e a variável β, chamada de taxa de decaimento, pode ser vista como uma espécie de ponderador da magnitude do decréscimo do período T, provocado pelo *resetting*. Este segundo termo só é ativado quando a unidade V-V considerada estiver sob o domínio do próximo pulso do oscilador acentual, $m+1$, e permanece ativo durante os dois primeiros ciclos V-V do grupo acentual que estiver sendo realizado.

O oscilador silábico é implementado matematicamente como uma função cossenóide levantada ($\frac{1}{2} \cdot [1 + \cos(2\pi t/T_0)]$) de amplitude unitária fixada arbitrariamente. Nos valores máximos desta função, a iteração da equação 1.3 fornece a atualização do período T.

$$T(n) = T(n-1) + \Delta T(n) \tag{1.3}$$

Na equação 1.3, $T(n)$ representa o valor do período do oscilador silábico durante a realização da unidade V-V$_n$, $T(n-1)$ é o valor do período na unidade V-V$_{n-1}$, isto é, a unidade anterior àquela considerada no momento e ΔT é o valor a ser acrescido (pode ser positivo ou negativo) ao atual valor do período do oscilador silábico, fornecido pela equação 1.2.

1.3 Acento secundário no quadro dinâmico

O modelo dinâmico do ritmo e o quadro teórico-metodológico mais amplo em que ele se insere oferecem algumas possibilidades para o tratamento do acento secundário. A escolha da melhor delas é uma questão empírica a ser decidida após o confronto os dados da descrição acústica. Como se disse na Introdução, os experimentos de produção relatados nesta tese servem como testes críticos para o modelo como formulado atualmente (Barbosa 2006 e 2007), uma vez que ele não prevê a existência de mais de um nível de proeminência em um mesmo grupo acentual, o que representaria um problema caso os dados apontassem a ocorrência sistemática de proeminências de base duracional distribuídas de forma binária, por exemplo.

Na hipótese dos dados fonéticos mostrarem a existência de alternâncias binárias nos contornos de duração de palavras polissilábicas, pelos menos dois cenários podem ser imaginados: mantem-se a formulação atual ou modifica-se o modelo pela adição de um oscilador que emulasse os efeitos do pé métrico binário.

Para manter a formulação atual do modelo a despeito de evidência favorável ao binarismo expresso pela duração, seria necessário postular a existência de tantos pulsos do oscilador acentual em uma palavra quantos fossem os acentos secundários. Supondo, a título de exemplo, que na palavra *possibilidade* no sintagma "A possibilidade" houvesse, de acordo com análises que supõem um binarismo estrito, dois acento secundários, nas sílabas *po* e *bi*, além de um acento frasal na sílaba tônica, o modelo sem modificações acomodaria este fato por meio da especificação de três batidas do oscilador acentual, resultando na segmentação prosódica[4]

[4] A barra horizontal marca a fronteira do agrupamento prosódico e o ponto a fronteira entre sílabas.

a po|ssi.bi|li.da|.de.

Esta solução, no entanto, não seria ideal. Ela contrariaria a intuição de alguns linguistas, que dizem que o acento secundário confere às sílabas que o portam um grau de força intermediário entre o do acento primário das tônicas (é com as tônicas lexicais que o acento frasal, no modelo, se alinha) e as sílabas átonas. Portanto, postular a atribuição, no domínio da palavra, de dois acentos de um mesmo nível hierárquico não parece uma boa saída. Seria possível salvar a manobra dando conta da relação hierárquica de força entre os hipotéticos "acentos frasais secundários" através de uma modulação apropriada de suas magnitures conseguida por meio da manipulação do parâmetro $i(m)$ da equação 1.2. Mesmo assim, essa solução não seria a mais desejável do ponto de vista da economia e da elegância porque seria necessário impor a relação entre as magnitudes dos acentos de forma arbitrária. Em comparação com a maneira mais natural como a sinergia entre parâmetros do modelo consegue reproduzir outros aspectos da organização rítmica da fala, os procedimentos descritos neste parágrafo não parecem a resposta mais adequada ao cenário em que exista evidência favorável ao binarismo implementado pela duração. Um exemplo desta sinergia entre os parâmetros do modelo pode ser observado na seção 4.7 onde são apresentados resultados de simulações realizadas com o modelo em sua atual formulação que mostram que a interação entre seus parâmetros enseja o surgimento de uma oscilação intermediária entre a do oscilador acentual e a do oscilador silábico na forma de uma proeminência inicial manifesta como o alongamento relativo do primeiro período do oscilador silábico. As simulações reproduzem com sucesso os resultados obtidos nos experimentos de produção reportados nos capítulos 3 e 4, em especial aqueles referentes ao parâmetro duração, em que se verifica a ocorrência de alongamentos da primeira unidade V-V de grupos acentuais em início absoluto de enunciado.

A solução para acomodar uma possível manifestação alternante de proeminências por meio de uma modificação do modelo consistiria em acrescentar um terceiro oscilador aos dois existentes. O adicional oscilador "podal", coextensivo ao pé métrico binário, atuaria em conjunto com o oscilador frasal no processo de indução do oscilador silábico. Seu papel seria provocar, no contorno de duração exponencialmente ascendente produzido pela ação de indução do oscilador frasal sobre o silábico, *resets* do oscilador silábico a intervalos regulares de forma a criar um padrão de alternância no contorno. O oscilador podal precisaria ser ajustado para que seu período mantivesse com o período do oscilador acentual uma relação harmônica de modo que se garantisse o aninhamento adequado das batidas do oscilador podal nas batidas do oscilador acentual. A relação entre os dois também poderia ser ajustada de modo a simular pés de aridades diferentes da binária. Do ponto de vista da parcimônia, quanto menor o número de osciladores necessários para simular o comportamento dos sujeitos humanos melhor. Uma formulação do modelo com três osciladores estaria, de acordo com esse critério, em desvantagem em relação à formulação atual, mais econômica. A introdução de um novo oscilador acrescentaria complexidade ao modelo pela adição de novos termos para expressar a força de acoplamento entre o novo oscilador e os outros dois previamente existentes. Portanto essa possibilidade só deveria ser usada caso os dados não deixassem nenhuma outra saída.

Os dois cenários discutidos anteriormente não esgotam as possibilidades de tratamento que os dados da descrição acústica podem sugerir. Além da duração, outros correlatos acústicos podem servir para a expressão de proeminências acentuais secundárias. Neste caso, não se

colocaria a princípio a questão de mudar a presente formulação do modelo de produção do ritmo, já que ele ainda não faz previsões para outros parâmetros. De qualquer forma, é interessante investigar quais as possibilidades que o quadro dinâmico em que o modelo de ritmo se insere pode apresentar para a acomodação de outros parâmetros acústicos além da duração.

A redução vocálica é uma possível candidata a parâmetro acústico de proeminências secundárias. Caso haja diferenciações sistemáticas na qualidade vocálica, tanto no sentido de uma redução das átonas quanto de um reforço das secundariamente acentuadas, este padrão poderia ser implementado por meio da especificação de alofones via gestos reduzidos, i.e., gestos vocálicos com menor magnitude relativa ou mais sobrepostos aos gestos consonantais vicinais. Para lidar com contrastes do tipo das alofonias, o modelo da fonologia articulatória, como exposto em Browman e Goldstein (1992), prevê duas situações: a de diferentes graus de sobreposição e magnitude relativa dos gestos ou a especificação de gestos diferentes para cada alofone. A proposição de um contraste entre gestos vocálicos plenos e reduzidos poderia acomodar tanto propostas em que os acentos secundários são motivados pela estrutura métrica mais global da frase quanto propostas em que o acento secundário é de natureza lexical. O recurso a este tipo de contraste no contexto da fonologia articulatória é encontrado, por exemplo, em Albano (2001), que propõe uma explicação calcada em alofones vocálicos reduzidos para os padrões de acento lexical do PB. Fornecendo evidência favorável à presença de informação fonética detalhada no léxico, Oliveira (2003), em estudo sobre a percepção da alofonia da vogal /a/ em PB, apresenta dados experimentais em favor da existência de alofonias lexicais categóricas ao mostrar que alofones artificialmente inseridos em posições impróprias demoram mais tempo para serem distinguidos em palavras do que em logatomas.

1.4 Integrando produção e percepção do ritmo

Como foi dito na Introdução da tese, no quadro teórico mais geral em que o modelo dinâmico do ritmo está inserido procura-se trabalhar de forma a não separar de maneira radical a produção e a percepção da fala. Desta forma, é natural que se questione as possibilidades que uma abordagem baseada em osciladores acoplados e em mecanismos como a indução podem oferecer para um tratamento integrado da produção e percepção do ritmo da fala. Barbosa (2006, pp. 440–442) delineia os traços gerais de uma proposta de abordagem integrada, que será discutida nesta seção.

Uma das tarefas em que um falante se engaja quando ouve um enunciado é identificar, com maior ou menor precisão, os pontos no enunciado que o interlocutor pretendeu tornar proeminentes, pois em geral estes pontos contribuem com informação linguística relevante para o processamento da frase (entre outras coisas, proeminências podem indicar quebras sintáticas que ajudem no estabelecimento da estrutura sintática ou colocar em foco um determinado trecho em oposição a outro). Esta tarefa de identificação é, no fundo, um trabalho de recuperação e reconstrução, uma vez que o ouvinte, não tendo acesso direto à representação linguística do enunciado assim como formulada por seu interlocutor, deve proceder usando o sinal de fala, seu conhecimento léxico-gramatical e elementos do contexto comunicativo como fontes de informação.

A tradução desta tarefa para os termos do modelo de produção do ritmo pode ser formulada

da seguinte forma: a tarefa dos ouvintes é tomar o sinal acústico e inferir os momentos em que ocorrem batidas do oscilador acentual, que podem ser em parte recuperadas pelo grau de proeminência percebido naquelas posições. O sinal acústico permite o acesso ao produto final do processo de acoplamento entre os osciladores acentual e silábico que ocorre durante a produção do enunciado. Os três quadros superiores na figura 1.2 representam a situação descrita ao longo do parágrafo. O primeiro quadro indica a localização dos pulsos do oscilador acentual e o segundo mostra os períodos do oscilador silábico sendo modificados pelo processo de indução deste pelo oscilador acentual. O terceiro quadro mostra o sinal acústico como um reflexo do resultado final do processo de acoplamento[5].

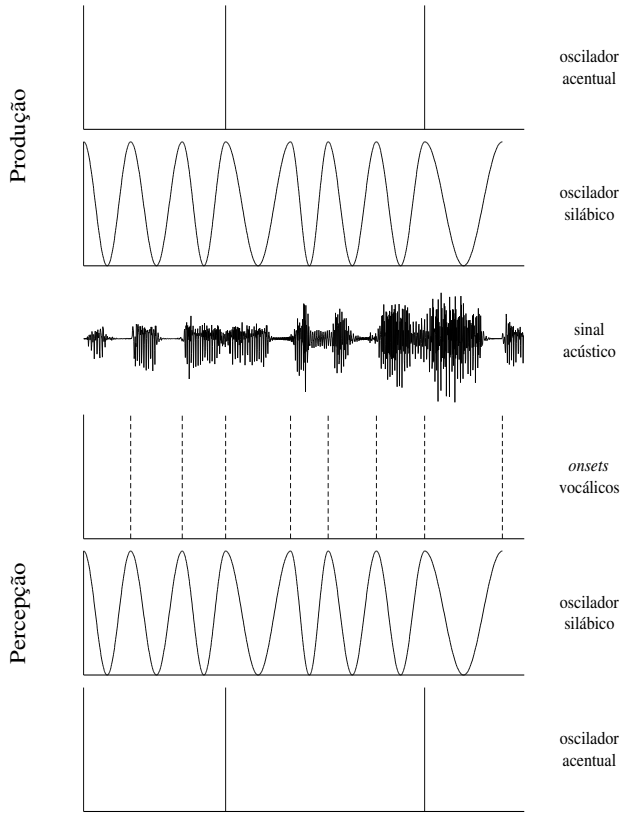


Figura 1.2: Representação esquemática da integração da percepção no modelo dinâmico do ritmo

Os três quadros inferiores mostram de forma simplificada como o ouvinte poderia fazer a recuperação dos momentos de ocorrência das batidas do oscilador acentual no enunciado a ser percebido. A detecção dos *onsets* vocálicos (marcados por linhas verticais traçejadas no

[5]Embora a figura possa sugerir isso, não se deve confundir o período do oscilador silábico induzido com a duração acústica dos segmentos. Conforme dito na seção 1.1 e representado na figura 1.1, a duração predita pelo modelo é produto da interação entre período do oscilador silábico e da pauta gestual.

quarto quadro) permitiria recuperar de uma forma aproximada os períodos do oscilador silábico do falante. Essa informação serviria como estímulo externo para um processo de indução do oscilador silábico interno do ouvinte, representado no quinto quadro da figura 1.2. Um mecanismo como esse, que permite o acoplamento entre estímulos externos e um oscilador perceptual interno é proposto por McAuley (1995). O acoplamento ajudaria o ouvinte a estimar a força do acoplamento entre os osciladores acentual e silábico e o período do oscilador silábico em repouso, respectivamente os parâmetros w_0 e T_0 do modelo de produção. Fazendo estas estimativas, o ouvinte poderia, monitorando momento a momento a duração dos intervalos entre *onsets* vocálicos, estabelecer pontos em torno dos quais é possível ter ocorrido um *reset* na atividade do oscilador silábico do falante, isto é, pontos em que a influência do oscilador acentual sobre o silábico é máxima seguidos de decaimento, retorno do período do oscilador silábico a valores próximos a T_0. Estes eventos servem para o ouvinte estimar, com alguma precisão, o momento de ocorrência da batida do oscilador acentual, representados no sexto quadro da figura 1.2.

O acoplamento do oscilador acentual com níveis linguísticos superiores (indicado na figura 1.1), um parâmetro que o ouvinte também deve deve ser capaz de estimar à medida que ganha familiaridade com seu interlocutor, também contribui para a estimativa feita pelo ouvinte do intervalo em que a ocorrência do pulso do acento frasal é mais provável. Detalhes fonéticos finos de duração segmental que são resultado da interação da prosódia com os segmentos também funcionariam como pistas adicionais no processo de inferência da posição do acento frasal.

Do ponto de vista da parcimônia, a proposta para a integração da produção e da percepção do ritmo da fala é interessante pela economia. Como a percepção, na proposta, pode ser vista como uma operação especular da produção, os mesmos recursos e princípios atuam nos dois lados e são ora mobilizados para a produção ora para a percepção conforme se esteja na posição de falante ou ouvinte.

Nesta tese, a proposta de uma abordagem integradora da percepção e da produção do ritmo não será desenvolvida nos seus pormenores. O que será investigada é a plausibilidade de uma das suposições assumidas pela proposta, que é a de que os falantes são sensíveis às mudanças sucessivas na duração dos intervalos definidos pelos onsets vocálicos ou unidades V-V. O capítulo 5 descreverá dois experimentos que buscam evidência comportamental para uma influência da duração dos intervalos intervocálicos em uma tarefa que envolve a percepção da fala. Esta evidência é importante porque a implementação da abordagem esboçada nesta seção só faz sentido como um modelo do processamento humano do ritmo se as operações e processos que ele postula tem plausibilidade. Os dois experimentos descritos no capítulo 5 buscarão evidência de que a percepção da sucessão dos intervalos intervocálicos é de algum modo afetada por sua estruturação temporal.

São necessários experimentos adicionais que respaldem a proposta de uma abordagem integradora da produção e da percepção de forma mais completa. Além de mais experimentos que consolidem o achado dos experimentos já realizados, que são evidência favorável do uso das unidades V-V como unidade básica da percepção do ritmo, são necessários experimentos que mostrem ser possível que os sujeitos façam, a partir da percepção dos *onsets* vocálicos, estimativas do período do oscilador silábico em repouso e da força de acoplamento entre os osciladores acentual e silábico. Seria necessário também fazer simulações com o modelo de McAuley ou uma versão modificada dele para testar seu desempenho na situação em que os

estímulos que causam a indução de período do oscilador silábico tem as características típicas produzidas pelo modelo de produção (no trabalho de McAuley 1995 os estímulos usados para induzir o oscilador interno não tem a estrutura complexa da fala). Além disso, é necessária a formulação matemática do mecanismo computacional responsável por gerar, a partir dos períodos do oscilador silábico induzido do ouvinte, previsões de ocorrências de batidas do oscilador acentual.

Capítulo 2

O acento secundário

O objetivo deste capítulo é apresentar o tema do acento secundário (doravante AS). Serão passadas em revista ideias e propostas de autores de diversas afiliações teóricas que de alguma forma têm relação com o tema. Esse panorama deve servir para mostrar, por um lado, como certos aspectos do tratamento do AS persistem ao longo do tempo e são incorporados em diversas tendências teóricas, formando linhas de continuidade. Por outro lado, há momentos em que as diversas definições do mesmo termo, acento secundário, nem sempre apontam de forma unívoca para os mesmos fenômenos. A resenha privilegiará mais as descrições do acento secundário apresentadas nos trabalhos do que a avaliação das propostas de análise.

A seção 2.1 apresenta algumas das abordagens tradicionais sobre o AS presentes nas literaturas portuguesa e brasileira. A seção 2.2 fala sobre como o trabalho de Mattoso Câmara Jr. pode iluminar a questão dos diversos graus de acentuação no âmbito da frase e como eles se relacionam com o acento na palavra. A seção 2.3 discute o tratamento dado ao tema pela fonologia pós-gerativa, em sua vertente métrica-autossegmental e na teoria da otimalidade. Nestas três seções, um traço comum à maior parte dos trabalhos ali apresentados é o uso de critérios puramente impressionísticos, baseados apenas na intuição do analista ou na outiva, para a constatação da existência e da localização dos acentos secundários. Mesmo as análises fonológicas contemporâneas partem de uma base evidencial baseada nas atribuições impressionísticas herdadas de uma tradição de cunho filológico[1].

A seção 2.4 apresenta alguns dos principais estudos experimentais sobre o AS em português brasieliro (doravante PB) que antecedem o presente e discute sumariamente seus resultados.

Como parte do esforço para apresentar um panorama e apontar convergências e dissensos na literatura sobre o AS, a seção 2.5 traz uma revisão, ainda que limitada, da pesquisa sobre o tema em outras línguas românicas. As línguas sobre as quais se encontrou material julgado relevante são o espanhol (2.5.1), o italiano (2.5.2) e o francês (2.5.3). Em função do interesse do presente trabalho em investigar intrumentalmente a natureza do fenômeno, privilegiaram-se as referências que de alguma forma procuraram correlatos acústicos do AS naquelas línguas. Para uma revisão biliográfica bastante completa, que descreve o fenômeno em um grande número de línguas de famílias linguísticas diversas, ver Castelo (2004).

[1]Para uma extensa crítica ao uso não-controlado de dados baseados na intuição linguística de analistas e falantes nativos, ver Schütze (1996).

2.1 Abordagens tradicionais

Gonçalves Viana (1973[1892]), ao expor a pronúncia do português europeu, afirma, tratando do tema da acentuação, que as proeminências na palavra e na frase podem ser apreciadas observando-se o grau de proeminência relativa das vogais, por ser este o elemento que em seu juízo melhor as caracteriza. Quanto à presença ou não de proeminência, distingue as vogais em dois tipos: tônicas e átonas. Subdivide as primeiras em: (a) "plenamente acentuadas", no caso em que apenas uma vogal sobressai sobre as demais. Os exemplos do autor são *rápido* e *fará* e (b) "primariamente e secundariamente acentuadas", quando há mais de uma vogal proeminente na palavra e entre as duas se estabelece uma relação de acento "principal" e "subordinado". Exemplo do autor (na grafia original): *ràpidaménte*[2] (note-se aqui que se trata de palavra bimorfêmica).

Said Ali (1908 [1895]) foi o primeiro a chamar a atenção para a existência no PB de "graus de atonicidade", expressão pela qual designava sua percepção de que entre as pré-tônicas das palavras polissilábicas nem todas são marcadas *na produção* pelo mesmo grau de atonicidade. Segundo exemplos do autor, em palavras como *dignamente* ou *arbitrado*, ouve-se um acento secundário em *di* e *ar*, respectivamente, o que leva o autor a postular três graus de acentuação relativa entre as sílabas de um polissílabo – o acento principal, o secundário e a ausência de acento. Ainda segundo o autor, "a *impressão* (grifo meu) geral é que as sílabas secundárias são alternativamente fracas e semifortes", padrão que seria mais facilmente perceptível em palavras com número maior de sílabas pré-tônicas, como "misericordioso" (exemplo do autor).

Antenor Nascentes (1960 [1922]), em suas considerações sobre a acentuação dos vocábulos em PB, refere-se à existência, em palavras de mais de três sílabas, do que chama de "acentos secundários". As palavras (na grafia original) *imperador*, *ràpidamente* e *generosidade* são oferecidas como exemplos e a localização dos AS é indicada pelo autor:

> No primeiro, além do acento principal da sílada *dor*, existe o secundário da sílaba *pe*; no segundo, há o acento principal da sílaba *men* e o acento secundário da sílaba *ra*; no terceiro, o acento principal da sílaba *da* e os acentos secundários das sílabas *ge* e *ro*. O vocábulo mais extenso da língua portuguesa é inconstitucionalìssimamente, que tem três acentos secundários (*cons*, *tu* e *li*). (pp. 33–34)

Percebe-se nestas considerações de Nascentes critérios ora semelhantes ora diferentes aos formulados por Said Ali. No caso de *imperador* e *generosidade*, Nascentes parece seguir o mesmo critério de Ali, uma atribuição alternante e binária (sílaba forte seguida de fraca) de acentos secundários nas pré-tônicas. Entretanto, discorda deste padrão nos casos de *ràpidamente* e *inconstitucionalìssimamente*. No primeiro, a atribuição parece ser motivada pelo processo morfológico que deriva "rapidamente" a partir de "rápido", o que sugere um acento secundário na palavra derivada como manifestação residual do acento lexical da palavra que serve de base para a derivação. No caso de *inconstitucionalìssimamente*, o caso parece mais complexo, pois se o AS em "li" pode ser explicado como um resíduo do acento principal da palavra "constitucional", os AS em "cons" e "tu" parecem ser gerados por algum tipo de regra de alternância binária, mas com as posições ajustadas para evitar encontros acentuais.

[2]Nos exemplos desta seção respeitou-se sempre a gafia empregada pelo autor.

Tomados como um todo, os exemplos arrolados pelos autores permitem formular hipóteses sobre possíveis regras para atribuição dos acentos secundários mas não explicações substanciais para o fenômeno.

2.2 Abordagem estruturalista: Mattoso Câmara Jr.

A proposta de Mattoso Câmara Jr. (1970a, 1970b e 1976) para a análise do acento nos vocábulos em PB reconhece a gradação da ocorrência da proeminência silábica e propõe uma notação convencional que associa o número 3 às vogais das silábas tônicas, 1 às pré-tônicas e 0 às pós-tônicas. Não reconhece, entretanto, a ocorrência de proeminências intermediárias entre 1 e 3, a não ser no grupo de força, definido como "emissão de fala contínua, sem pausa intercorrente" (Mattoso Câmara Jr. 1976, p. 36). Reconhecendo a tendência do PB de reservar a proeminência principal dos constituintes prosódicos para sua borda direita, o autor atribui o grau 2 de proeminência para os acentos lexicais subordinados, isto é, os anteriores ao último acento principal do grupo.

Seguem abaixo as pautas prosódicas dos vocábulos "celebridade" e "célebre idade" onde se indica a força de cada vogal segundo a convenção do autor.

1 1 1 3 0
selebridade

2 0 1 3 0
selebridade

O que estes exemplos sugerem é a não pertinência de AS de base lexical, i.e., que diferenciem as pré-tônicas entre si (uma vez que na notação do autor todas as pré-tônicas recebem o mesmo grau de proeminência), mas sim uma espécie de subordinação dos acentos lexicais presentes no mesmo grupo de força ao acento mais à direita.

2.3 Abordagens gerativistas

Os trabalhos apresentados nesta seção analisam o fenômenos do AS a partir da perspectiva da fonologia gerativa. O primeiro trabalho apresentado trata a questão a partir da teoria métrica (Liberman e Prince 1977 e Prince 1983) e os outros dois adotam a teoria da otimalidade (Prince e Smolensky 1993). A apresentação que será feita aqui se concentrará apenas na descrição dos padrões de ocorrência do AS que os autores assumem e que procuram explicar em suas análises, sem fazer avaliações técnicas sobre as mesmas.

Em sua análise do fenômeno em PB, Collischonn (1994) assume uma descrição segundo a qual o AS se manifesta por meio de dois padrões. O primeiro deles é observado em palavras com um número par de sílabas pré-tônicas, condição na qual os AS ocorrem sempre de maneira binária, a começar da primeira sílaba (exemplos[3]: *àlmofáda*, *pròbabìlidáde* e *ìrrespónsabílidáde*). O segundo padrão, que pode ser observado no caso da palavra possuir um número ímpar de

[3]Todos os exemplo nesta seção foram retirados dos trabalhos comentados. Os AS são marcados pelo acento grave (`) e acentos primários pelo acento agudo (´).

sílabas pré-tônicas, é criado por um processo de aplicação opcional chamado pela autora de "recuo do acento secundário" que dá origem a agrupamentos ternários, isto é, um AS seguindo de duas átonas. Entre outros, a autora lista os seguintes exemplos deste padrão: *àbacaxí* e *àcondicíonaménto*. A autora considera as sílabas pré-tônicas o único domínio de aplicação e descarta a existência de AS de base lexical.

Sandalo e colaboradores (2006) também assumem uma descrição em que o AS se manifesta segundo um padrão primordialmente binário, reconhecendo também o fenômeno que Collischonn (1994) chama de recuo do AS[4]. Sandalo e colaboradores, no entanto, mencionam que a análise acústica do *corpus* analisado em seu trabalho revela a ocorrência de apagamentos vocálicos ("vowel deletion" no original) em palavras contendo um número ímpar de sílabas pré-tônicas que, segundo a análise dos autores, seriam motivados pela tendência da língua a formar agrupamentos binários. Um dos exemplos mencionados pelos autores é o apagamento da vogal [i] na palavra *modernização*. Em sua análise, após o apagamento ritmicamente motivado da vogal [i], a palavra seria analisada em dois constituintes métricos (*mò dern*)(*za çã´o*).

Lee (2002) em sua análise assume a existência de dois tipos de AS. Além de reconhecer o padrão que pode ser descrito como binário e é assumido também por Collischonn (2006) e Sandalo e colaboradores (2006), o autor indica a existência de um padrão de atribuição de AS que ele chama de lexical. Este último padrão só poderia ser explicado, segundo o autor, se a estrutura interna da palavra fosse levada em conta no momento da atribuição dos AS. Segundo Lee (2002), o a AS da palavra *jacàr*[ɛ]*zínho* seria um exemplo de AS atribuído de forma lexical.

2.4 Estudos experimentais em PB

A pesquisa sobre os correlatos acústicos do acento lexical em PB mostra (Massini 1991) que a duração é seu principal correlato acústico e que seus efeitos se estendem não apenas sobre a vogal, mas sobre toda a sílaba. Barbosa (1996), por sua vez, contribui para o tema explorando as possibilidades da duração como correlato tanto do acento lexical quanto do acento frasal. Quando os dados normalizados de duração segmental são agrupados em sílabas sobressai a contribuição do acento lexical. Quando o agrupamento se dá em torno das unidades V-V[5], revela-se o domínio do acento/acentuação frasal. Em relação a correlatos acústicos do AS em PB a literatura é menos conclusiva. Não há consensos muito estabelecidos sobre qual ou quais parâmetros são mais relevantes nem sobre a relação do AS com o acento lexical e com o acento frasal.

O trabalho de Gama Rossi (1998) confirma a tendência, aludida pela autora, a serem escassas na literatura as evidências de manifestação, em dados de duração, da alegada alternância binária, basilar para a postulação do algoritmo de atribuição do acento secundário. Analisando dados de duração, frequência fundamental (f_0) e qualidade vocálica em polissílabos paroxítonos derivados (como *macacada* e *palhaçada*), constata: (a) inexistência de evidência de alternância para o parâmetro duração, pois os contornos de duração normalizada são sempre ascendentes da esquerda para a direita entre as pré-tônicas das palavras presentes em seu *corpus*; (b) ocor-

[4]Segundo os autores, nos exemplos 1–6 (p. 98), a posição dos AS foi determinada pela consulta a falantes nativos, embora os detalhes desta consulta não sejam apresentados.

[5]Unidade compreendendo os fones entre dois *onsets* de vogal consecutivos (ver Barbosa 1994 e 1996).

rência em apenas alguns casos de diferença significativa entre os valores de f_0 na transição da primeira para a segunda pré-tônica e (c) redução estatisticamente significativa da segunda vogal pré-tônica da palavra "palhaçada" e maior abertura da segunda vogal pré-tônica na palavra "macacada"[6]. Os resultados da análise de f_0 e configuração formântica oferecem suporte para a postulação de uma verdadeira realização de proeminência secundária, embora bastante variável, uma vez que ela se realiza ora na forma de um *pitch accent* na primeira vogal pré-tônica, ora na mudança da qualidade vocálica da segunda vogal pré-tônica, que pode ainda ser no sentido da abertura ou do fechamento. No caso de "macacada", todas as variáveis apontam para a realização da proeminência gradativa culminando no acento frasal. Neste caso, a proeminência é marcada não por alguma característica específica (duração, valor de f_0, intensidade etc) de uma determinada unidade na cadeia da fala (a vogal, a sílaba etc) em relação às demais, mas pela mudança progressiva de determinados parâmetros acústicos ao longo de cada grupo acentual. Esse fato descaracteriza o fenômeno do acento secundário assim como proposto pelas abordagens expostas acima.

Moraes (2003) apresenta resultados de um estudo perceptual e uma análise de medidas acústicas sobre o tema. Em seu trabalho, cinco conjuntos com palavras de composição segmental semelhante e relacionadas morfológica e semanticamente (ex. secular/ secularizo/ secularizar/ secularização), cujo número de sílabas pré-tônicas variava entre duas e cinco, foram gravados por quatro falantes. No estudo perceptual, dez sujeitos ouviram os conjuntos gravados e pediu-se que eles apontassem dentre as sílabas pré-tônicas quais eram secundariamente acentuadas. Observou-se uma variabilidade bastante grande nas respostas, embora o autor identifique dois padrões principais: alternância binária e proeminência inicial. Quanto aos dados relativos aos parâmetros acústicos, os sujeitos não empregaram, na produção das frases lidas por eles, estratégias consistentes de uso da duração silábica, frequência fundamental e intensidade para assinalar possíveis proeminências secundárias, frequentemente utilizando combinações deles. O trabalho ressente-se, no entanto, do pouco controle sobre a influência da duração intrínseca dos segmentos (que poderia ser resolvido pelo uso de algum tipo de normalização) e da não utilização de testes de inferência estatística para dar suporte à análise tanto dos resultados do teste percetual quanto das medidas acústicas. A falta de normalização não representa um problema sério para as comparações paradigmáticas, pois a composição segmental das sílabas pré-tônicas é a mesma em cada conjunto de palavras. Ficam prejudicadas as comparações sintagmáticas intraenunciado.

Arantes e Barbosa (2002), em um estudo piloto, verificaram o comportamento dos padrões de duração em uma palavra para a qual se esperaria encontrar alternância, o polissílabo paroxítono "macacada", inserindo-o num conjunto de frases-veículo que permitiram controlar o contexto fonético (a palavra testada foi a mesma), o número de sílabas (o número de sílabas por frase divergia no máximo em um), o acento frasal (testou-se a influência da distância da sílaba proeminente no nível da frase em relação ao início da frase). Os dados de duração foram agrupados em sílabas e unidades V-V (o que já representou uma novidade em relação aos trabalhos experimentais anteriores) e as análises estatísticas aplicadas mostraram a não existência de uma alternância marcada pelo parâmetro duração, uma vez que não foi encontrada diferença signi-

[6]Este padrão de comportamento, segundo Barbosa (1996), é o esperado para a realização do acento frasal em PB

ficativa entre as durações (normalizadas por *z-score*) médias da primeira e da segunda sílaba da palavra. A duração das unidades V-V, por sua vez, conformou-se em um padrão ascendente até a realização do acento frasal, sem que se verificasse qualquer padrão alternante. Além disso, evidência estatística mostrou que o alongamento duracional das posições pré-tônicas, tanto considerando as sílabas quanto as unidades V-V, é influenciado pela posição da unidade na cadeia sintagmática do enunciado, de forma mais significativa do que pela posição da unidade considerada relativamente à unidade de maior força no grupo acentual (acento frasal). Os mesmos testes mostraram, por outro lado, que as posições lexicalmente acentuadas recebem um grau de alongamento significativamente diferente conforme estejam ou não nas proximidades do ponto culminante do acento frasal. No primeiro caso, o alongamento é maior.

2.5 O fenômeno em outras línguas românicas

2.5.1 Espanhol

Análise fonológica

A contribuição original da proposta de análise de Rocca (1986) para o AS em espanhol é tratá-lo como parte do componente pós-lexical da fonologia. Segundo sua proposta, o AS é atribuído de forma independente do acento primário ou lexical, cujo domínio de aplicação é a palavra prosódica. Seu domínio seria o grupo fônico[7], o mesmo em que atua a regra que atribui o acento sintagmático (*main phrasal stress*, no original). Nisso difere de propostas como a de Collischonn para o PB (cf. seção 2.3) e de Vogel e Scalise (cf. seção 2.5.2) para o italiano, que não levam em conta nenhum domínio acima da palavra. Roca apresenta dados e análises para argumentar em favor da necessidade da regra de AS ser alocada no componente pós-lexical em função de interação com regras claramente pós-lexicais, como as regras de atribuição de acento sintagmático e de fusão silábica (*syllable merger*, no original).

Análises anteriores do AS em espanhol e outras línguas românicas como o italiano, analisam-no como o resultado da aplicação de regras de acento particulares a cada língua. O problema destas análises, segundo Roca, é que elas não fazem distinção clara entre os algoritmos de atribuição de acento primário e secundário e implicitamente consideram o AS como parte das regras do componente lexical.

As análises do autor, que não serão reproduzidas aqui, tentam trazer evidência para a inconsistência resultante da aplicação de regras de atribuição de AS no componente lexical e argumentar em favor da necessidade de grades acentuais construídas sobre estruturas a que já se atribuiu acento primário ou lexical.

O caminho percorrido pelo autor sugere uma dicotomia entre acentos atribuídos pelo componente lexical (chamados por ele de acentos lexicais) e acentos pós-lexicais, atribuídos pelo componente pós-lexical. Esta divisão traria, segundo sua opinião, as vantagens de (i) liberar o acento lexical das pressões de conformidade a um ritmo prosódico baseado em alternâncias e (ii) liberar os acentos pós-lexicais das eventuais idiossincrasias da determinação do acento lexical (como, p.e., sensibilidade a peso silábico e fatores morfológicos). Note-se a possibilidade

[7] Cuja definição é semelhante à do grupo de força de Câmara Jr. (cf. seção 2.2, p. 15.)

de contato entre esta proposta de Roca de diferenciação do acento lexical e dos pós-lexicais e a visão dinamicista (cf. seção 1.3), na qual o acento lexical vem especificado no léxico via presença ou ausência de gestos reduzidos e os acentos frasais assinalados pelo modelo de ritmo.

Dados acústicos e de percepção

Prieto e van Santen (1996) procuraram evidências acústicas para o AS em espanhol. Não assumem de antemão qualquer teoria ou modelo que suponha sua existência, enunciando, a partir de uma revisão da bibliografia que consideram relevante sobre o tema, três hipóteses de trabalho: (a) hipótese rítmica, segundo a qual os AS apresentam-se em alternância binária; (b) hipótese não-rítmica, que prevê a realização do AS como uma proeminência atribuída no início de palavras com mais de duas sílabas pré-tônicas e (c) hipótese da não existência, prevendo a não-existência de proeminências secundárias na palavra em estilos de fala não enfáticos.

O material de análise utilizado por eles é um *corpus* de 550 palavras com até quatro sílabas pré-tônicas, selecionadas de maneira a atender a combinação de fatores que afetam a duração das vogais em espanhol, considerados relevantes pelos autores. As palavras foram inseridas na frase-veículo "El murmura *palavra-chave* de nuevo", lidas por um locutor mexicano que produziu três repetições de cada frase. Os parâmetros acústicos duração vocálica, f_0 e intensidade (estes dois últimos apenas na forma de inspeção de contornos particulares) foram analisados. A análise estatística posterior mostrou que o parâmetro acústico duração vocálica não mostrava evidência de uma alternância binária. A vogal das sílabas pré-tônicas, malgrado sua distância em relação à tônica, apresentaram duração média estatisticamente indistinguível entre si qualquer que fosse o número de pré-tônicas da palavra considerada. Os autores reportam, entretanto, a ocorrência de movimentos descendentes da f_0 e picos de amplitude mais elevados associados à vogal da pré-tônica inicial quando há pelo menos duas delas. Estes indícios de um possível marcação de proeminência inicial não são consubstanciados por uma análise estatística para confirmar a suspeita. Apenas alguns gráficos mostrando contornos particulares de f_0 e de intensidade são mostrados a título de ilustração. A força deste achado é em parte obscurecida em razão das poucas evidências de sistematicidade apresentadas. O uso de algum tipo de técnica de estilização do contorno de f_0 das frases produzidas pelo locutor ou de um sistema de transcrição de entoação poderia contribuir para atestar de forma mais convincente a aplicação generalizada do fenômeno.

Um experimento de percepção foi realizado no qual dois falantes nativos do espanhol[8] escutaram um subconjunto de 100 frases do *corpus* do estudo de produção e foram solicitados a indicar, na palavra-chave, a(s) sílaba(s) em que eles percebessem proeminências secundárias e também a classificar estas proeminências em termos de sua força de acordo com uma escala de cinco valores (não fica claro em relação a que outro evento este julgamento deveria ser feito). Um dos sujeitos indicou a sílaba inicial como portadora de proeminência secundária em todos os casos apresentados. O outro sujeito teve o mesmo comportamento em 93% dos casos. Os autores não reportam nenhum dado referente ao julgamento dos sujeitos quanto à força desta proeminência inicial. A validade de um teste como esse é em parte comprometida pelo fato de a tarefa solicitada apresentar alguns inconvenientes. Um deles é o de ser de natureza acen-

[8]Não há informação sobre qual era a variedade de espanhol falada pelos participantes. Um deles é autor do trabalho em questão (Pilar Prieto, de nacionalidade espanhola).

tuadamente metalinguística, o que pode obrigar os sujeitos a mobilizar recursos e estratégias de percepção diferentes daqueles que ele emprega na atividade ordinária de percepção da fala. Além disso, e mais grave, a amostra do experimento de percepção foi extremamente reduzida, um dos sujeitos participantes não era ingênuo em relação aos objetivos do experimento e ainda era falante nativa da mesma variedade do espanhol do locutor do experimento.

Os dados do experimento de produção e de percepção convergem no sentido de apontar a primeira sílaba como portadora de uma proeminência, possivelmente sinalizada em conjunto pelos parâmetros acústicos f_0 e intensidade. Como este padrão de proeminência secundária única e em posição inicial ocorre independentemente do número de sílabas pré-tônicas da palavra-chave, os autores interpretam os resultados dos experimentos como evidência contra uma atribuição binária e alternante para as proeminências secundárias, o que confirmaria a hipótese (b), AS não-rítmico, citada no primeiro parágrafo da seção.

Dados acústicos e articulatórios

Scharf et. al. (1995) e Dogil (1999)[9], reconhecendo o relativo insucesso da procura de correlatos acústicos para os AS postulados por diversas análises fonológicas do espanhol, buscam evidência para a existência de diferenças na articulação de sílabas a que se pode atribuir, segundo as análises fonológicas, diferentes graus de acentuação. A justificativa para esta estratégia baseia-se em dados (de Jong et. al. 1993) que indicam existir diferenças articulatórias entre sílabas átonas e tônicas em inglês. Os movimentos articulatórios envolvidos na produção das tônicas, segundo este trabalho, apresentam maior velocidade e amplitude do que os medidos para as átonas.

Em seu experimento, os autores escolheram como variável independente as diferentes condições acentuais impostas sobre a mesma sílaba em um conjunto de palavras-chave. Os autores escolheram estudar possíveis diferenças articulatórias provocadas por diferents graus de acento na sílaba /ti/ das seguintes palavras:

a) Constan***ti***no

b) constan**ti**no*ple*ño

c) constan**ti**nople*ar*

As sílabas lexicalmente acentuadas estão destacadas em itálico. A sílaba de interesse no experimento está destacada em negrito. No item a) a sílaba /ti/ é a tônica lexical, em b) ela porta, por hipótese, o AS e em c) ela é átona. Estas palavras foram inseridas na frase-veículo "No he dicho *palavra-chave* jamas." e cada versão dela foi lida, a partir de sua apresentação visual, oito vezes por um falante nativo de espanhol peninsular[10].

A articulação da sílaba /ti/, nas palavras-chave do *corpus* é realizada principalmente pela ponta da língua, que faz um movimento aproximadamente circular que começa na região alveolar durante o movimento de oclusão do [t], desce durante a realização do [i] e sobe novamente

[9]Dogil (1999) é uma republicação, na forma de capítulo de livro, de Scharf et. al. (1995), que é um trabalho curto publicado nos anais de um congresso. O experimento relatado nas duas referências é o mesmo.

[10]Segundo informação presente no texto, o sujeito gravado é um dos autores em Scharf et. al. (1995), nomeadamente Iggy Roca.

para a oclusão alveolar do [n] da próxima sílaba. Um articulógrafo registrou, a uma taxa de amostragem de 200 Hz, as coordenadas bidimensionais, no eixo sagital, de um sensor instalado na ponta da língua do sujeito[11]. Sensores foram ainda instalados nas regiões medial e posterior da língua e também nos lábios superiores e inferiores, embora as informações sobre o movimento destas partes não seja crucial para a caracterização da articulação em estudo. Quatro eventos acústicos — meio da oclusão de [t], ataque, ponto médio e *offset* da vogal [i] — foram escolhidos como marcos para a coleta dos valores de deslocamento horizontal máximo no sentido anterior, deslocamento descendente máximo, deslocamento horizontal posterior máximo e deslocamento vertical ascendente máximo, respectivamente. A partir destes dados foi possível recompor o deslocamento total no plano horizontal (abreviado por dx) e no plano vertical (dy). A soma $dxy = dx + dy$ foi usada como aproximação da magnitude do movimento e o quociente $dq = dy/dx$, medida da relação entre as componentes vertical e horizontal do movimento, como aproximação da forma (*shape*, no original) do movimento da ponta da língua. Além de dxy e dq, mediu-se também a duração da sílaba /ti/.

A análise estatística dos dados de duração tanto da vogal quanto da sílaba mostram que há uma diferença significativa (ANOVA com $\alpha = 5\%$) entre a condição de acento lexical (maior duração) e as demais condições em conjunto. Essa diferença não existe na comparação entre as sílabas (e vogais) átonas e com suposto acento secundário.

Quanto aos dados articulatórios, uma análise de componentes principais revelou que os fatores magnitude (dxy) e forma (dq) do movimento são relevantes para diferenciar as categorias de acento entre si. Testes *post hoc* indicaram as seguintes diferenças significativas: 1) a média da variável dxy para a condição de acento lexical é diferente das demais condições — as sílabas tônicas apresentaram as maiores amplitudes nas dimensões horizontal e vertical e 2) as médias da variável dq na condição de atonicidade e de suposto AS são significativamente diferentes entre si — as átonas tendem a apresentar valores maiores na dimensão vertical, enquanto nas sílabas com AS eles estão na dimensão horizontal. Os autores argumentam que essa diferenciação entre a condição átona e de AS pode não ser muito robusta, uma vez que a composição segmental das palavras-chave, não sendo exatamente igual nas três condições acentuais, pode ter provocado essa variação na composição da forma do movimento por conta de algum efeito coarticulatório não controlado. Além disso, se uma das hipóteses iniciais fosse que as sílabas com suposto AS devessem ter características articulatórias intermediárias entre a tonicidade máxima e a atonicidade, o resultado obtido é de certa forma inesperado. Uma das razões para afirmar isso é que um dos efeitos produzidos pela maior magnitude dos movimentos articulatórios característica das sílabas tônicas é aumentar a intensidade acútisca da vogal tônica em função do maior deslocamento para baixo da mandíbula e língua em conjunto. Ora, seria de esperar, portanto, que as sílabas com AS apresentassem, na relação entre as componentes vertical e horizontal do movimento, valores de maior vulto exatamente na vertical como estratégia para conseguir um reforço acústico (e possivelmente perceptual). O resultado não foi, portanto, exatamente o esperado no caso de uma influência robusta do AS na articulação, já que no experimento relatado pelos autores a influência do suposto AS é apenas um reforço da dimensão horizontal do movimento de articulação da sílaba /ti/ que mesmo assim pode ser um artefato introduzido pela diferença na composição segmental das diferentes palavras-chave.

[11] Isto significa que a cada 5 ms a posição do sensor era registrada.

2.5.2 Italiano

Análise fonológica e dados acústicos

Bertinetto (1976 e 1981)[12], resenhando de forma abrangente a literatura linguística italiana que faz referência ao conceito desde a década de 40 do século passado, assinala a polivalência e ambiguidade do conceito de acento secundário. Entre os nove principais autores citados, nota-se uma pluralidade de usos e definições do conceito que torna impossível qualquer tentativa de definição unívoca do termo. A razão para tanto é o fato de a natureza dos fenômenos agrupados sob a mesma designação ser em verdade muito diferente. O autor, em um esforço sistematizador, propõe que a diversidade de empregos do termo AS com respeito ao italiano pode ser esquematicamente reduzida a três principais acepções:

(a) proeminência acentual automática que aparece em sequências de pelo menos duas sílabas átonas;

(b) acento condicionado lexicalmente, fonologicamente pertinente e subordinado ao acento primário em palavras compostas;

(c) acento hierarquicamente condicionado, subordinado ao acento primário em um sintagma (de forma que se pode falar em acentos n-ários, a depender do sintagma).

O fenômeno a que se refere o item (a) não deve receber a denominação de AS na opinião do autor, uma vez que este tipo de proeminência é, segundo ele, distinto dos acentos observáveis no âmbito do que ele chama de "fonologia da palavra". Sua existência se deve à tendência, presente na fala, de organizar a distribuição de proeminências na enunciação de maneira a conformar-se a princípios gerais, como euritmia ou minimização de esforço. Tendências deste tipo, nota o autor, estão presentes em diversas línguas além do italiano, transformadas em restrições, como por exemplo a de evitar sequências longas de sílabas não acentuadas. Bertinetto cita autores que notam a atuação de princípios semelhantes em línguas como inglês, holandês, alemão e tcheco. As proeminências geradas por este mecanismo devem, propõe o autor, ser chamadas de acentos rítmicos (AR), para deixar clara a distinção entre estes acentos e aqueles cujo domínio é a palavra. Esta distinção é importante, uma vez que mesmo na palavra, quando isoladamente pronunciada, pode haver a atribuição de AR, que não devem ser confundidos com acentos de origem não-rítmica.

Quanto à acepção (b), o autor toma de empréstimo critérios de Garde (1968) para afirmar que o AS não deve ter sua posição dedutível a partir da posição do acento principal, sendo, ao contrário, determinada lexicalmente em unidades acentuais distintas que juntam-se por meio de algum processo de formação de palavra. Tal situação dá ensejo ao aparecimento de pares mínimos, opostos pela colocação dos acentos primário e secundário no interior das unidades adjungidas. Línguas em que a formação de compostos por aglutinação é um processo produtivo na formação de palavras (como é o caso do alemão, citado por ele) atestam um número abundante de pares mínimos desta natureza. O italiano, por outro lado, salvo exemplos

[12] O trabalho de 1981 é basicamente a reimpressão, com algumas modificações, do trabalho de 1976 ma forma de um capítulo de livro.

cunhados por Lepschy[13] (considerados artificiais e de escassa representatividade no léxico por Bertinetto e outros autores italianos), não conhece este tipo de AS. Ainda a título de contraste, ele nota que no alemão os morfemas são unidades acentuais autônomas, enquanto no italiano as evidências fazem crer, segundo Bertinetto, a autonomia é verificada na palavra.

O tipo de acento descrito em (c) pode ser observado, segundo o autor (apoiado em Martinet 1961), em línguas que manifestam tendência a criar escalas acentuais de grau mais elevado do que dois, i.e., apresentam acentos terciários, quaternários etc. Neste caso, a diferença entre graus de acentuação não diz respeito a hierarquias estabelecidas entre elementos (sílabas, vogais, etc) no interior de uma mesma palavra, mas sim entre as sílabas acentuadas de um conjunto de tamanho variável de palavras num mesmo constituinte prosódico. O alemão e o inglês são as línguas, segundo Bertinetto, em que a discussão sobre graus acentuais mais vicejou, em função da relevância semântica e do papel funcional que a hierarquia acentual assume nestas duas línguas.

O tipo de material mais discutido pelos defensores da ideia de graus de acentuação (citados e brevemente resenhados pelo autor) são sintagmas do tipo *determinante + determinado* em que a relação semântica de determinação não é suficientemente marcada do ponto de vista morfossintático. Dois exemplos citados pelo autor são *kitchen towel rack*, que pode significar 'cabide para [toalha de cozinha]' ou '[cabide de toalha] para cozinha' e ainda *black bird's nest*, cujos sentidos possíveis são '[ninho preto] de pássaro' e 'ninho de [pássaro preto]'. Nestes dois exemplos, as duas possíveis interpretações semânticas são conseguidas modificando-se o grau de proeminência, numa escala de três valores possíveis, das sílabas tônicas das palavras nos sintagmas.

O autor argumenta em favor da não relevância deste tipo de acento, denominado por ele acento subordinado (ASb), no italiano, uma vez que naquela língua não há a necessidade de marcações adicionais além das morfossintáticas em sintagmas do tipo em questão. Ele faz referência a certas situações no italiano que poderiam ser analisadas como casos de subordinação de acento, como sintagmas do tipo *demonstrativo + substantivo* (p.e. 'quell'uomo' ou 'questo giorno') nos quais o acento do demonstrativo pode aparecer enfraquecido. A interpretação preferida por Betinetto, entretanto, é a de enfraquecimento do acento por cliticização e não subordinação no sentido definido acima.

Experimento Além de discutir o fenômeno do ponto de vista qualitativo, Bertinetto montou um *corpus* para investigar do ponto de vista experimental algumas das asserções mais recorrentes e importantes a respeito do fenômeno das proeminências secundárias. O *corpus* contém palavras e sintagmas de três tipos:

(a) pares do tipo *portar veli* vs. *portarveli*;

(b) palavras compostas, como *televisione*, *pomodoro* e *selvaticamente*;

(c) polissílabos monomorfêmicos, tais como *vagabondare*, *domomite* e *opportunità*.

[13] Entre outros, o suposto par mínimo *affáscinauómini* 'aquilo que fascina os homens' vs. *affascìnauómini* 'aquilo que reúne os homens'.

Metodologia As palavras foram gravadas em duas situações: pronunciadas em isolamento e em contexto, i. e. inseridas em uma frase-veículo (diferente para cada palavra) por falantes do dialeto piemontês. Os sintagmas foram gravados apenas em contexto. Os parâmetros acústicos duração vocálica, f_0 e intensidade[14] foram extraídos nas palavras do *corpus* e analisados.

Bertinetto não fez análises estatísticas para os parâmetros, apresenta tabelas com valores médios obtidos para cada palavra, o que mitiga a possibilidade de generalização dos resultados para além dos itens lexicais escolhidos no experimento. Para avaliar a pertinência das diferenças de duração entre as médias de duração obtidas para as vogais em cada palavra ele adota como critério para comparar a diferença entre os valores de duração de vogais adjacentes um limiar de 200 ms que ele toma de empréstimo a Lehiste (1970). Isso é de certa forma um problema, uma vez que a utilização de um critério de percepção para caracterizar a produção do AS pode constituir uma mistura indesejável de critérios. Além disso, a não utilização de normalização dos dados de duração torna os resultados menos confiáveis, pois a influência da duração intrínseca dos segmentos não foi minimamente controlada.

Resultados Apesar das críticas aos procedimentos metodológicos, alguns resultados indicativos são interessantes. Segundo o autor, o comportamento dos parâmetros f_0 e intensidade não permite qualquer tipo de generalização, sendo apenas a duração o parâmetro que em seu *corpus* permite traçar indicações para a expressão fonética de proeminências sejam primárias ou de outra ordem.

Quanto aos pares do tipo exemplificado no item (a), *portar veli* e *portarveli*, nos quais alguns autores esperariam encontrar um AS no segundo caso (*portarveli*), os dados mostram que os dois termos do par diferem entre si por meio de diferentes esquemas de duração. A diferenciação atestada, todavia, não sugere, segundo o autor, a necessidade de postular mais do que dois níveis de acentuação (elementos acentuados vs. não-acentuados). No caso de *porta bagagli* e *portabagagli*, a produção do sintagma apresentou alguma variabilidade entre os informantes quanto a presença de um relevo importante na primeira sílaba de *porta*. Em *portabagagli*, palavra pronunciada em isolamento apresenta (três casos em quatro) algum relevo sobre a primeira sílaba, enquanto em contexto a proporção cai para dois casos em quatro.

No caso dos compostos, muitas análises propõem a existência de um AS na sílaba originariamente acentuada no primeiro elemento do composto. Nas palavras *selvaticamente* e *televisione* não há qualquer indício disso. Nas demais, a presença de alguma evidência interpretável (sempre segundo seus critérios) como AS no local onde as análises prognosticam é bastante variável entre os informantes gravados. Além disto, a magnitude dos candidatos a AS diminui quando a palavra está inserida em contexto.

No conjunto de palavras não-derivadas, a presença de possíveis AS é bastante variável, manifestando-se mais consistentemente em algumas palavras para todos os informantes e em outras sua manifestção é variável entre os informantes. O autor nota que no caso da pronúncia em contexto há uma tendência geral de enfraquecimento das proeminências não ligadas ao acento lexical.

[14]Não há referência explícita sobre o método de extração destes dois últimos parâmetros nem sobre se os valores reportados dizem respeito ao valor no ponto médio das vogais ou representam o valor médio no intervalo de duração das vogais.

A conclusão a que chega Bertinetto é a de não pertinência da ideia de AS na fonologia italiana. O tipo de proeminência que aparece nos resultados pode ser melhor racionalizada em sua opinião como a manifestação de acentos rítmicos em função de sua variabilidade quanto aos lugares onde aparece (menos variável no caso de alguns compostos e mais livre em outros tipos de palavra), sensibilidade a fatores como o contexto frástico e possivelmente taxa de elocução.

Análise fonológica

Ao contrário de Bertinetto, que põe em questão a relevância de um grau de acentuação intermediário entre atonicidade e tonicidade no nível da palavra, Vogel e Scalise (1982) assumem a existência de AS na língua italiana. Reconhecem a dificuldade de identificar precisamente as sílabas portadoras de AS numa palavra tanto a partir de critérios objetivos (os autores mostram conhecimento do trabalho de Bertinetto e dos resultados pouco claros em favor do AS) quanto por meio da consulta à intuição dos falantes nativos, já que diferentemente de línguas como holandês e inglês, dizem eles, o italiano não tem pistas conspícuas de inacentuabilidade, como vogais reduzidas pré-tônicas.

A análise dos autores indica que em italiano a posição do(s) AS na palavra isolada pode ser predita, desde que se leve em conta não apenas critérios fonológicos, como peso silábico e a contagem de sílabas pré-tônicas, mas também informação sobre a composição morfológica da palavra em questão. Vogel e Scalise postulam três restrições básicas para a distribuição de AS em palavras não compostas: acento inicial, evitar encontro acentual (*stress clash*) e evitar sequências com mais de duas sílabas átonas. Além disso, acrescentam quatro regras de reajustamento (regra contra encontro, regra de reversão de acento, regra de acento inicial e regra de inserção de acento) que, devidamente ordenadas, produzem uma distribuição de AS em conformidade com a intuição dos falantes consultados pelos autores.

Deve-se notar uma convergência desta análise em relação à analise de Bertinetto, que é a importância dada às proeminências iniciais. Nos dados experimentais de Bertinetto, a vogal da sílaba inicial, principalmente das palavras não compostas, apresenta-se em muitos casos comparativamente alongada em relação à vogal da sílaba adjacente, principalmente na pronúncia da palavra em isolamento. Este alongamento é entendido pelo autor como uma "oscilação acentual", devida à presença de um acento rítmico, cuja presença é mais forte na palavra pronunciada em isolamento. Pode-se imaginar que a intuição do falante nativo sobre o padrão acentual de uma dada palavra da língua seja informada, entre outras coisas, pela internalização de uma pronúncia prototípica desta palavra. Pode-se ainda imaginar que essa pronúncia internalizada na memória de longo termo do falante seja muito próxima de instanciações efetivas daquela mesma palavra em contexto isolado. Uma vez que é comum produzir um alongamento inicial que afeta a vogal da sílaba inicial em italiano seria de esperar que a intuição dos falantes nativos, em sendo as hipóteses acima enunciadas verdadeiras, apontasse a presença de acentos em início de palavra.

2.5.3 Francês

Hoskins (1994) discute a possível relação entre fenômenos ligados à situação de encontro acentual e a existência de acento secundário em francês. A proposta de representação do

acento em francês adotada pelo autor, inspirada na fonologia prosódica (Nespor e Vogel 1986), é sensível à existência de encontro acentual no nível da palavra fonológica (ou grupo clítico, a depender da estrutura considerada) e estas ocorrências devem, portanto, ser resolvidas de algum modo. Hoskins propõe-se a procurar, por meio do exame de dados experimentais, possíveis correlatos acústicos da resolução do encontro e investigar se esta resolução tem alguma relação com a existência, na língua francesa, de acentos secundários em duas estruturas particulares: os substantivos compostos e os sintagmas nominais atributivos.

A presença e a maneira de atribuir AS em francês são assuntos controversos, havendo por um lado quem negue sua existência e, por outro, aqueles que, mesmo admitindo sua relevância, discordam a respeito da forma de sua manifestação. Entre estes últimos há quem argumente em favor de AS atribuídos de forma alternante à esquerda do acento primário (Tranel 1986) e outros (Dechaine 1990) que identificam dois tipos de moldes entoacionais[15]: (a) padrão iâmbico, no qual os AS ocorrem alternadamente a contar do início da palavra e (b) arco acentual, no qual a primeira sílaba recebe o AS e a última o primário. Mazzola (1992) sugere, em consonância com a ideia de arco acentual, que o padrão de acento mais comum em substantivos compostos e sintagmas nominais é SWS e SWWS[16].

Para tanto, o autor montou um *corpus* de cinco conjuntos de substantivos compostos e três de sintagmas nominais atributivos formando paradigmas como os exemplificados na tabela 2.1.

Cada paradigma segue um esquema em que o primeiro elemento tem duas sílabas (*Marie* em I e *dernier* em II) e o segundo elemento pode ter uma, duas ou três sílabas (linhas *a*, *b* e *c*, respectivamente). O segundo elemento funciona como a variável independente do experimento, criando ambientes em que a resolução do encontro pela aplicação da reversão iâmbica é obrigatória (Ia e IIa), opcional[17] (Ib e IIb) e não aplicável (Ic e IIc). As hipóteses de trabalho elaboradas pelo autor fazem previsões sobre a possível influência da situação de encontro sobre a configuração acústica das duas sílabas do primeiro elemento de cada substantivo ou sintagma. Estas hipóteses fazem as seguintes previsões: (H1) nenhuma proeminência nas duas sílabas no caso de não haver AS; (H2) atribuição de AS na primeira sílaba se o total de sílabas do composto ou sintagma for ímpar e na segunda sílaba se o número total for par; (H3a) se a situação de encontro acentual não for tolerada deve haver uma reversão iâmbica (*iambic reversal*), i. e., atribuição de AS na primeira sílaba nos casos semelhantes aos ilustrados em Ia e IIa e AS na segunda sílaba nos demais casos e (H3b) a situação de encontro pode ainda ser resolvida por apagamento de um dos acentos, resultando em nenhuma proeminência observável nas duas sílabas iniciais nas condições Ia e IIa e AS na segunda sílaba nos demais casos.

Cinco sujeitos (sendo quatro provenientes da região de Paris) produziram dez repetições de cada condição experimental, sendo seis delas escolhidas para análise posterior. Cada estímulo foi inserido numa frase-veículo declarativa em posição posterior ao verbo principal. Os parâmetros acústicos investigados foram a duração silábica (valores brutos, sem nenhum procedimento de normalização), a amplitude das vogais, a f_0 medida no ponto médio das vogais e a taxa

[15] Não fica claro se o uso da palavra *entoacional*, aqui, indica algum compromisso sobre o parâmetro responsável pela caracterização fonética destes moldes ser a f_0, como faria imaginar o uso convencional do termo.

[16] O autor utiliza os símbolos W e S para representar sílabas fracas (átonas) ou fortes (acentuadas em algum grau), respectivamente

[17] O autor, seguindo a análise de Mazzola (1992), assume que neste contexto há encontro acentual (ver Hoskins 1994, exemplo 12, p. 39), contrariamente à análise clássica de Liberman e Prince (1977) e Prince (1983).

	I – sub. compostos	**II** – SN atrib.
a)	*Marie-Jo*	*dernier mot*
b)	*Marie-Joëlle*	*dernier moment*
c)	*Marie-Joséphine*	*dernier monastère*

Tabela 2.1: Esquema dos estímulos do experimento de Hoskins (1994).

de mudança da f_0 da primeira para a segunda sílaba. Os dados resultantes passaram por análises de variância (ANOVA) para apurar possíveis diferenças significativas entre as médias amostrais. Os dados de cada sujeito e de cada um dos dezoito tipos de estímulos (8 expressões $\times$ 3 formas por paradigma) foram analisados separadamente. A variável independente é o número de sílabas no segundo elemento do composto ou do sintagma e a variável dependente os parâmetros acústicos aferidos na primeira e segunda sílabas dos compostos e sintgmas.

As ANOVA mostraram não haver diferença significativa entre as médias da primeira e da segunda sílabas para os parâmetros duração silábica, amplitude das vogais e valor da f_0 no ponto médio das vogais. O único padrão interpretável, segundo o autor, foi aquele apontado pelo parâmetro taxa de mudança da f_0 entre a primeira e a segunda sílaba de cada composto ou sintagma, verificado em quatrodos cinco sujeitos do experimento. A condição *a* (cf. Tabela 2.1), tem os menores valores médios de taxa de mudança de f_0 (doravante f_0'). A condição *b* tem valores intermediários: 41% das ocorrências são estatisticamente iguais aos valores médios da condição *a* e 39% equivalem estatisticamente aos valores médios de *c*; no restante dos casos, o valor é distinto das duas outras condições. A condição *c*, por sua vez, tem os valores médios mais altos de f_0'.

Os resultados são muito pouco conclusivos em função de algumas escolhas metodológicas do autor, como a não utilização de procedimentos de normalização dos dados de duração, cuja análise pode ter sido bastante comprometida pelo fator duração intrínseca dos fones e também pela opção de analisar separadamente os dados de cada sujeito e de cada estímulo separadamente, o que torna a generalização dos resultados mais difícil. Apesar das eventuais limitações dos resultados, eles parecem favorecer a hipótese H3b formulada pelo autor e descrita em parágrafo anterior. Esta hipótese prevê o apagamento do acento do primeiro elemento do composto ou do sintagma no caso da condição *a* e atribuição de AS na segunda sílaba do primeiro elemento nas demais condições. A interpretação de apagamento do acento no caso da condição *a* parece razoável em função dos baixos valores de f_0' apresentados por esta condição, cuja variação de f_0 quase desprezível autoriza que se diga que entre as duas sílabas consideradas nenhuma é mais proeminente que a outra. A respeito da atribuição de AS na segunda sílaba, pode-se dizer que os valores de f_0' apresentados pela condição *c* podem em princípio ser vistos como a sinalização de uma culminância acentual localizada na segunda sílaba, o que corroboraria a hipótese H3b. Os resultados gradientes observados na condição *b*, no entanto, são mais dificilmente acomodados pelas previsões da hipótese H3b, já que ora aproximam-se do padrão de apagamento de acento previsto para a condição *a* ora aproximam-se do padrão previsto para a condição *c*. Essa variabilidade não é prevista ou explicável segundo os princípios adotados pelo autor, nem mesmo no caso da alternância ser declarada opcional, situação em que deveria haver 50% de chance da condição *b* manifestar-se acusticamente de

forma idêntica à condição *a* ou *c*. Não é o caso que isso ocorra, uma vez que em 20% dos casos o valor médio de f_0' na condição *b* distingue-se igualmente das duas outras. O fato de haver uma possível sinalização de acento por meio da mudança mais brusca da f_0 na condição *c* e em alguns casos na condição *b* na segunda sílaba do primeiro elemento de cada composto ou sintagma configuraria um caso de acento secundário no sentido, apontado por Bertinetto (cf. na seção 2.5.2, p. 22), de acento hierarquicamente condicionado, subordinado ao acento primário em um sintagma. O acento da primeira palavra nos substantivos compostos ou SN atributivos (*Marie* e *dernier*, respectivamente nos exemplos da Tabela 2.1) manifestam-se, mas sua magnitude está subordinada à distância entre o acento do primeiro e do segundo elementos. Nos casos em que a distância é pequena ou mínima (condição *a*) há o que o autor chama de "apagamento" do primeiro acento. Nas demais condições, a distância sendo maior, a manifestação do primeiro acento pode ser maior, mas sujeita a gradiência.

2.6 Discussão

O panorama apresentado aqui sugere que os resultados de grande parte dos experimentos dedicados a investigar correlatos de proeminências secundárias não consegue fornecer evidência positiva para a confirmação das hipóteses elaboradas pelos investigadores. Aparentemente, parece haver um certo descompasso entre a sofisticação de algumas dentre as diversas propostas que postulam a existência do AS e a pouca evidência conseguida através de experimentos formulados para testar as hipóteses que subjazem às análises teóricas.

É possível que estes resultados pouco satisfatórios sejam fruto de opções metodológicas que conduzem a limitações observacionais. Em parte, essas opções são tributárias de algumas convicções arraigadas no campo, como a ideia segundo a qual os efeitos fonéticos da atribuição do acento limitam-se à sílaba (ou mesmo apenas à vogal) que o recebe, o que leva muitos autores a circunscrever a observação dos parâmetros investigados apenas ao núcleo da sílaba considerada acentuada. O experimento de Bertinetto (seção 2.5.2 é um exemplo desta prática. Mesmo aqueles que observam uma unidade maior do que a rima quase sempre limitam-se a expandir seu universo observacional para no máximo as sílabas da palavra investigada. Isso ocorre mesmo quando supõe-se que o domínio de aplicação dos supostos AS pode ser maior do que a palavra, como é o caso do experimento de Prieto e van Santen (seção 2.5.1), que cita como uma de suas fontes o trabalho de Roca (seção 2.5.1), cujo principal argumento é mostrar que o AS tem como domínio de aplicação o que ele chama de grupo fônico e mesmo assim limitam-se a estudar as sílabas de uma palavra.

Além do problema da limitação observacional, há outra ordem de inconvenientes extensível à maioria praticamente absoluta dos trabalhos resenhados. O primeiro deles é a falta generalizada do uso de algum procedimento de normalização dos valores de duração apurados experimentalmente. Quando se deseja verificar se a duração pode ser considerada um bom correlato da organização prosódica, o pesquisador deve tentar eliminar ao máximo a influência de outras fontes de variabilidade que não aquelas de interesse. A não utilização da normalização introduz como fonte de variação a duração intrínseca[18], e obscurece a possível influência do mecanismo de acentuação na manifestação do parâmetro de interesse.

[18]Para uma explicação um pouco mais detalhada sobre duração intrínseca, cf. anexo A

Parte da dificuldade encontrada na interpretação dos dados comportamentais produzidos pelos experimentos pode estar na pouca difusão da prática de propor modelos matemáticos e/ou computacionais de produção da fala capazes de simular os padrões acústicos ou articulatórios que correspoderiam a determinadas características ou arranjos presentes nas representações linguísticas. O modelo de ritmo apresentado na seção 1.3 é um exemplo nesse sentido, na medida em que permite propor hipóteses sobre a manifestação do parâmetro acústico duração a partir de uma determinada descrição linguística. Pode-se, então, comparar os resultados de um experimento com a fala natural com aqueles previstos pelo modelo.

Capítulo 3

Estudo de produção I

O panorama traçado no capítulo anterior mostra que o conhecimento sobre a possível manifestação fonética de AS em PB e em outras línguas românicas ainda é pequeno e que algumas insuficiências no campo metodológico podem contribuir para esse cenário. O experimento relatado no presente capítulo é uma contribuição no sentido de iluminar a natureza do AS e investigar seus possíveis correlatos acústicos em PB.

O principal objetivo do experimento é obter uma descrição foneticamente bem controlada do acento secundário em PB, de modo que seja possível caracterizar quais são as circunstâncias em que se pode atestar a existência de proeminência secundária e como ela se traduz em termos de correlatos acústicos (duração, f_0 e configuração formântica). Em termos metodológicos, o experimento aprofunda a abordagem adotada no estudo piloto (seção 2.4, p. 17) controlando um maior número de variáveis.

Um outro objetivo do experimento é o de testar a adequação da atual formulação do modelo dinâmico de ritmo, investigando se ele pode simular os contornos duracionais de frases contendo palavras polissilábicas, situação que muitos acreditam dar ensejo à observação de proeminências secundárias. As possibilidades que o modelo do ritmo em particular e o quadro dinamicista em geral têm para dar conta do problema são descritas na seção 1.3. Dessa forma, os resultados obtidos no experimento deverão contribuir para a discussão a respeito das teorias dinamicistas de produção da fala.

3.1 Metodologia

3.1.1 Variáveis independentes

Para a realização do experimento foi montado um *corpus*, constituído de frases-veículo (ver lista 1 na página 33), nas quais foram inseridas palavras-chave (ver tabela 3.1). O *corpus* foi elaborado de maneira a controlar um certo número de variáveis linguísticas e compará-las de maneira ortogonal.

Num experimento com desenho de comparação ortogonal, cada fator controlado é variado sistematicamente ao longo de dimensões definidas pelo experimentador, enquanto os demais são mantidos constantes. Desta maneira, consegue-se fatorar a influência de cada um dos fatores

ou variáveis para a composição do comportamento geral do fenômeno estudado e ao mesmo tempo manter a condição *ceteris paribus*. A seguir, a lista de variáveis independentes:

- o número de posições na palavra-chave (PS). Como posição, deve-se entender tanto o número de sílabas quanto o de unidades V-V na palavra-chave. Sendo assim, um quadrissílabo tem quatro posições silábicas e cinco posições V-V. Numa frase-veículo como "*A patarata p*arece menor hoje.", as unidades V-V são delimitados da seguinte maneira no trecho destacado[1]: /a# p/$_1$ /at/$_2$ /ar/$_3$ /at/$_4$ /a# p/$_5$. Assim, o número de unidades V-V é sempre igual a $n+1$, sendo n o número de sílabas da palavra-chave. Nas análises estatísticas, cada posição é considerada um nível do fator PS. A adoção da unidade V-V como unidade de agrupamento da duração é justificada por Barbosa (1996), que mostra que a duração da sílaba é um índice da acentuação da palavra e que a duração das unidades V-V pode ser tomada como manifestação do grau de acentuação no nível do enunciado. Como pareceu importante controlar a influência deste último tipo de acentuação, sobre a manifestação do acento secundário, fez-se necessário observar um trecho maior do que a palavra-chave (inclui-se no trecho delimitado pelas unidades V-V, p. e., o artigo que precede a palavra-chave);

- a estrutura da sílaba inicial da palavra-chave (ES), que pode ser V ou CV;

- a natureza das vogais da rima silábica (VR);

- a distância entre a tônica da palavra-chave e a culminância do acento frasal (d_σ). Define-se a variável independente d_σ como o número de sílabas contadas no intervalo $]L\ F]$, em que L é a sílaba tônica da palavra-chave e F a sílaba em que se espera que ocorra a culminância do acento frasal.

Estas variáveis permitem controlar a influência de certos fatores na manifestação da acentuação secundária, permitindo isolar a hipotética contribuição específica do fenômeno em questão. A seguir, a adoção de cada uma das variáveis é justificada.

Se a ocorrência de proeminências secundárias no nível da palavra obedecer às previsões baseadas nas análises que postulam a alternância binária como tendência subjacente, então o *corpus* do experimento deve conter palavras-chave com duas ou mais sílabas pré-tônicas, de modo a criar-se o ambiente necessário para a manifestação da regra de alternância binária. Desse modo é possível verificar se ela ocorre e, em caso afirmativo, qual parâmetro ou conjunto de parâmetros pode ser candidato a correlato. Outra hipótese que é colocada em teste por meio do emprego de polissílabos do tipo especificado acima é aquela derivada do modelo dinâmico do ritmo. Ela afirma que o parâmetro duração provoca, por meio de um mecanismo de culminação, o aumento progressivo e contínuo da duração de unidades do tamanho da sílaba de modo a delimitar grupos acentuais compreendidos entre dois pontos de culminância.

A adoção da variável ES visa esclarecer a intuição de Mattoso Câmara Jr. (1970a e 1970b) segundo a qual sílabas sem ataque consonantal são menos propensas a receber proeminências.

[1] O símbolo # será usado para representar fronteira entre palavras

Para neutralizar a influência da duração intrínseca das diferentes vogais, procurou-se escolher palavras-chave nas quais pelo menos todas as sílabas pré-tônicas compartilhem a mesma vogal em suas rimas, constituindo esta a variável VR.

Escolheu-se, finalmente, verificar a influência da acentuação frasal sobre a realização das proeminências no grupo acentual, o que é conseguido variando-se a posição da culminância do acento frasal em relação à palavra-chave, afastando progressivamente aquela primeira desta última. Este é o papel que realiza a variável d_σ no *corpus*, conseguido pela adição de adjetivos à direita da palavra-chave nas sentenças-veículo.

3.1.2 Frases-veículo

(1) a. A *palavra-alvo* parece menor hoje. (d_σ= 0)
b. A *palavra-alvo* **ru**de parece menor. (d_σ= 2)
c. A *palavra-alvo* ru**ral** parece menor. (d_σ= 3 o[xítona])
d. A *palavra-alvo* bu**dis**ta parece menor. (d_σ= 3 p[aroxítona])
e. A *palavra-alvo* bico**lor** parece menor. (d_σ= 4)

Nas frases, o determinante sempre concorda em gênero e número com o alvo e o verbo e o predicativo do sujeito concordam em número com o alvo. No exemplo **1** estão em negrito as sílabas tônicas dos adjetivos. Nas situações em que existe um adjetivo aposto espera-se que a cabeça do grupo acentual seja deslocada para a sílaba tônica deste, de modo a aumentar o número de sílabas no grupo acentual. Quando a palavra-alvo é *patarata*, por exemplo, os grupos acentuais, delimitados pelo símbolo "|", podem apresentar ao menos as duas distribuições ilustradas em **2**.

(2) a. |A patarata bicolor| parece menor.
b. |A patarata| bicolor| parece menor.

No exemplo **2a** a palavra *patarata* e o adjetivo formam um só grupo acentual com 8 unidades V-V e no exemplo **2b** a palavra-alvo forma um grupo acentual com 5 unidades V-V e o adjetivo um outro grupo com 3 unidades V-V. No corpus de produção que tem sido analisado no âmbito do presente projeto verifica-se a ocorrência dos dois agrupamentos prosódicos. Pode-se observar, no entanto, que a palavra-alvo pode variar de posição relativamente à sua posição no interior do grupo acentual. No caso do agrupamento ilustrado em **2a** a palavra-alvo, especialmente sua sílaba tônica está em posição medial, em contraste com o que ocorre em **2b**, em que a tônica do alvo coincide com a posição de acento frasal. É importante lembrar, aqui, que o modelo dinâmico de produção do ritmo que tem sido adotado neste projeto, prevê que as unidades mais próximas à culminância do acento frasal são aquelas que mais sofrem consequências duracionais da sua implementação. É uma hipótese que foi em parte corroborada por etapas anteriores da presente pesquisa que grupos acentuais mais longos, como o exemplificado em **2a**, em contraste com grupos mais curtos, como o que se vê em **2b**, pode favorecer o surgimento de uma proeminência na borda esquerda dos grupos, se estes estão em posição de início absoluto de enunciado. Relatos deste efeito de proeminência inicial estão reportados em Barbosa, Arantes e Silveira (2004) e Arantes e Barbosa (2006), além de Barbosa (2006).

Deve-se notar ainda que, em geral, se acredita que quanto maior a cadeia de sílabas sem um acento lexical ou frasal maior seriam as chances de um acento secundário acontecer, em função de algum tipo de restrição ao número de sílabas átonas em sequência. A variabilidade no agrupamento prosódico permitida pela variável d_σ permite, assim, testar esta hipótese particular a respeito do AS.

3.1.3 Palavras-alvo

A tabela 3.1 lista as palavras selecionadas para o *corpus*. Para a escolha das palavras que atendessem a todos os parâmetros desejados, foram utilizados os mecanismos de busca da versão eletrônica dos dicionários *Aurélio* e *Houaiss*.

VR	PS	ES	frequência	palavra
a	2	V	0,2	*arataca*
		C	—	*patarata*
	3	V	—	*arapabaca*
		C	—	*jaratacaca*
i	2	V	—	*iridite*
		C	0,5	*p[i]riquito*
	3	C	—	*pirimidina*
	5	C	2,2	*dirigibilidade*
u	2	V	—	*urucaca*
		C	0,3	*pururuca*
	3	V	0,2	*urucubaca*
o	2	V	0,8	*horroroso*
		C	—	*gonococo*
	3	V	—	*hodofobia*
		C	2,5	*locomotiva*
	4	C	—	*colonoscopia*
e	2	V	—	*edesseno*

Tabela 3.1: Lista de palavras-chave do *corpus* do experimento de produção I. A frequência de uso é listada em ocorrências por milhão.

Foram escolhidas palavras que atendessem na medida do possível a dois critérios: (a) que fossem monomorfêmicas e (b) cujas sílabas pré-tônicas tivessem a mesma vogal em seu núcleo. O objetivo é isolar influências da composição morfológica na ocorrência de possíveis acentos secundários. A exigência da identidade entre as vogais das sílabas pré-tônicas foi adotada para que, numa mesma palavra, se pudesse neutralizar a influência da vogal na manifestação de possíveis acentos secundários.

Uma variável lexical para a qual não foi possível fazer uma variação sistemática foi a frequência de uso em função da dificuldade de conciliar o controle desta variável com a exigência

de monomorfismo e a de igualdade das vogais nas pré-tônicas. As palavras do *corpus* inicial são de baixa ou baixíssima frequência de ocorrência[2]. As frases-veículo formadas com as palavras-chave podem ser criticadas por sua pouca naturalidade semântica. Este inconveniente pode ser visto como um custo necessário para se conseguir condições de controle e comparabilidade que minimizem as fontes introdutoras de variabilidade nos parâmetros experimentais observados. Apesar dos inconvenientes, o sujeito não reportou nenhuma dificuldade para realizar a leitura das frases e exame posterior realizado no material gravado não encontrou nenhum traço de surpresa ou hesitação que pudesse ser atribuído à pouca naturalidade semântica das frases.

3.1.4 Procedimentos de gravação

Foram obtidas dez repetições lidas de cada frase-veículo, gravadas na cabine acusticamente tratada do Laboratório de Fonética e Psicolinguística (Lafape/IEL/Unicamp) por um locutor masculino adulto (em torno de vinte anos na época das gravações), falante nativo da cidade de Descalvado–SP. As gravações foram armazenadas em fitas DAT e posteriormente digitalizadas a uma taxa de amostragem de 22,05 kHz no programa CSL, modelo 4300 B, da Kay Elemetrics. A ordem de apresentação das frases foi aleatorizada a cada sessão de repetição.

O *corpus* definido para o projeto é composto por dezessete palavras-chave, cada uma delas sendo inserida em cinco frases-veículo, sendo cada combinação possível repetida dez vezes pelo sujeito informante, totalizando 850 frases gravadas e digitalizadas.

3.1.5 Variáveis dependentes

Duração

As durações foram delimitadas utilizando o programa Praat, a partir da comparação entre o gráfico de forma de onda sincronizado com o espectrograma, tendo-se utilizado tanto informação espectral quanto da forma de onda na delimitação. Delimitou-se cada sinal gravado em fones, separando vogais e consoantes, sendo os ditongos ou realizações ditongadas medidos como um intervalo acústico único. Em cada frase-veículo foi preciso delimitar a duração de 15 fones em média (este número varia em função do tamanho da palavra-chave e da sentença em que ela foi inserida), totalizando em torno de 12.750 delimitações. Estas unidades delimitadas foram agrupadas em três subtipos de variáveis: duração de rima, de sílaba fonológica e de unidade V-V.

- As 9 palavras de 4 sílabas resultaram em 2198 dados de duração de unidades V-V, 1760 dados de duração de sílabas e 1760 dados de duração das rimas.
- As 6 palavras de 5 sílabas resultaram em 1.698 dados de duração das unidades V-V, 1406 dados de duração de sílabas e 1406 dados de duração das rimas.
- A palavra de 6 sílabas (*colonoscopia*) resultou em 294 dados de duração das unidades V-V, 245 dados de duração das sílabas e 245 dados de duração das rimas.

[2]Há, inclusive, palavras que, apesar de dicionarizadas, não são atestadas sequer uma vez num *corpus* de uso do português brasileiro contendo mais de 200 milhões de ocorrências.

- A palavra de 7 sílabas (*dirigibilidade*) resultou em 392 dados de duração das unidades V-V, 343 dados de duração das sílabas e 343 dados de duração das rimas.

Os valores de duração em unidades de tempo (*ms*) foram normalizados utilizando a transformação *z-score*. Conferir o apêndice A.1 para a justificativa para o uso da transformação e para a fórmula matemática de sua implementação.

Frequência fundamental (f_0)

Uma das maneiras mais tradicionais de analisar o parâmetro acústico f_0 tem sido a extração do seu valor no ponto médio da duração acústica das vogais em trechos selecionados de enunciados. Esta foi uma das estratégias adotadas no presente experimento. Um *script* escrito para o programa Praat colheu os valores de f_0 (calculado por meio do método de auto-correlação, o mais indicado para o estudo da entoação) no ponto médio da duração das vogais da palavra-chave e também do adjetivo que seguia a palavra nos casos em que a variável d_σ é diferente de zero, bem como na vogal do artigo que precede a palavra-chave. Assim, no caso da frase-veículo "[O gonococo budista] parece menor", foram marcados oito pontos de extração da f_0 correspondendo ao ponto médio de cada vogal presente no constituinte delimitado pelos colchetes. A premissa segundo a qual a entoação deve organizar a variação da f_0 num escopo maior do que simplesmente a palavra-chave motivou a observação dos valores da f_0 nestas posições.

Apesar de tradicional, a abordagem da variação de f_0 por meio do estudo de valores pontuais apresenta algumas limitações que são discutidas no apêndice B. Desenvolvemos uma abordagem alternativa em que a variabilidade da curva de f_0 é reduzida aos seus picos e vales, tornando possível então contar o número de ocorrências desses eventos ao longo de uma palavra ou grupo acentual e estabelecer o alinhamento dos picos ou vales em relação às unidades V-V ou sílabas. O apêndice B descreve o algoritmo desta abordagem alternativa. As análises dos dados obtidos por meio da aplicação do algoritmo descrito no apêndice B avaliaram os efeitos do número de sílabas pré-tônicas sobre os padrões de distribuição de extremos altos (H) e ascendentes (R) ao longo das palavras-chave.

Configuração formântica das vogais

O parâmetro acústico intensidade contribui para a sensação perceptual de volume do som (*loudness*, em inglês). A intensidade dos sons da fala é subsidiária, em certa medida, da abertura mandibular, de modo que há uma certa correlação entre abaixamento de mandíbula e aumento da intensidade sonora irradiada. Se o abaixamento de mandíbula ajuda a incrementar a intensidade acústica de uma vogal e, consequentemente, a possibilidade de ser percebida como proeminente, então a presença de vogais produzidas com maior abaixamento relativo de mandíbula pode representar um possível correlato para proeminências secundárias na palavra. Na falta de dados articulatórios que comprovem a existência de abaixamentos diferenciais da mandíbula pode-se recorrer às inferências articulatórias que podem ser feitas a partir dos dados acústicos (Lindblom e Sundberg 1971). No caso do abaixamento de mandíbula, a pista acústica é o aumento no valor do primeiro formante (F_1).

Uma outra possibilidade para a configuração formântica desempenhar algum papel como correlato acústico de proeminências secundárias é a possível diferenciação na qualidade das

vogais. Há dois cenários imagináveis neste caso: (a) as vogais átonas poderiam sofrer algum processo de redução, verificável por meio de mudanças nos dois primeiros formantes, no sentido de centralização e/ou fechamento e (b) as vogais secundariamente acentuadas poderiam sofrer algum tipo de reforço sem que as átonas sejam necessariamente reduzidas.

Escolheu-se para análise da estrutura formântica das vogais a palavra *dirigibilidade* por ser a palavra com maior número de sílabas pré-tônicas no *corpus* do experimento, partindo do princípio segundo o qual quanto maior for o número de pré-tônicas numa palavra maior serão as chances de aparecer proeminências secundárias. Foram também analisadas as palavras *patarata* e *jarataca* para cobrir cadeias com 2 e 3 pré-tônicas. A escolha destas palavras em particular foi motivada pelo fato de, diferentemente de *dirigibilidade*, elas tem a mesma vogal em todas as sílabas, inclusive o artigo definido que a precede na frase-veículo. Uma outra razão é que a vogal /a/ é a que mais se adequa à metodologia de extração de ênfase espectral que será descrita na seção 3.1.5).

Para a extração dos valores de frequência formântica, utilizou-se o programa Praat para calcular o espectro FFT no ponto médio de cada vogal por meio de um *script*. Em seguida, outro *script* calculava o espectro LPC no mesmo ponto e os dois eram sobrepostos para verificação. Caso não houvesse congruência entre ambos, o espectro LPC era recalculado mudando a ordem de filtro. Este procedimento baseou-se em heurísticas indicadas por Vallabha e Tuller (2002), que sugerem um ajuste individual da ordem de filtro para cada vogal ao invés de um ajuste para todo um *corpus*, mesmo que produzido por um mesmo falante. Foram analisadas em torno de 850 vogais.

O abaixamento da mandíbula é uma variável que está relacionada com a duração acústica, já que excursões mais amplas da mandíbula na dimensão vertical levam mais tempo para serem completadas, tornando mais longo o som produzido durante a excursão. Serão feitos testes de correlação entre os valores de F_1 e de duração para estabelecer o grau de independência entre os dois parâmetros.

Ênfase espectral

A ênfase espectral é uma medida da distribuição relativa da energia em diferentes faixas do espectro de um determinado fone. A importância desta medida para o estudo do acento deriva da evidência de que quando um som é produzido com mais esforço vocal por parte do falante, a região das altas frequências do espctro desse som tendem a receber um reforço de energia em compração com o mesmo som produzido com menos esforço vocálico. Segmentos acústicos produzidos com mais esforço vocal tendem a ser mais proeminentes em relação as segmentos produzidos com menos esforço laríngeo. Há evidência (ver Sluijter e van Heuven 1996) de que, em algumas línguas, sílabas tônicas em posição de foco prosódico tendem a ser produzidas com mais esforço vocálico. Portanto, é possível imaginar que a ênfase espectral pode ser um possível correlato para o AS.

Estimativas para a ênfase espectral foram obtidas seguindo-se aproximadamente a metodologia encontrada em Sluijter e van Heuven (1996), uma das fontes mais citadas pela literatura que aborda o tema. Naquele artigo, os autores escolheram palavras e logatomas com a vogal /a/ em seus núcleos silábicos e essa decisão afetou os detalhes da extração da ênfase espectral. A escolha do subconjunto de palavras-chave feita no presente trabalho foi influenciada pela

escolha feita no artigo citado.

Em Sluijter e van Heuven (1996) a escolha da vogal /a/ como alvo da análise é justificada pela possibilidade de isolar f_0, F_1, F_2 e F_3+F_4 em quatro bandas contíguas e sem superposição, de forma que se torna possível identificar de forma mais ou menos precisa qual formante ou grupo de formantes recebe ênfase em cada cirscunstância. As quatro bandas definidas em Sluijter e van Heuven (1996) são: B1: 0 a 0.5 kHz, B2: 0.5 a 1 kHz, B3: 1 a 2 kHz e B4: 2 a 4 kHz.

Uma função de pesquisa do programa Praat que calcula diferenças de energia entre bandas definidas pelo usuário foi empregada para obter-se a medida de ênfase espectral. Essa função foi aplicada aos espectros de FFT gerados para a análise de formantes descrita na seção anterior. Foram calculadas as diferenças entre cada uma das bandas consecutivas descritas no parágrafo anterior. Os rótulos B2-1, B3-2 e B4-3 serão doravante utilizados para fazer referência às diferenças (expressas em dB) entre B2 e B1, B3 e B2 e, por fim, B4 e B3.

Serão realizados testes de correlação entre os valores de ênfase espectral e duração para estimar o grau de independência destas duas variáveis.

3.1.6 Análise estatística

A significância da influência das variáveis independentes sobre a média amostral das variáveis dependentes foi testada por meio de análises de variância (ANOVA), adotando-se sempre um nível de significância (α) de 5%. No caso de comparações múltiplas recorreu-se ao teste *post hoc* de Scheffé para verificar o nível de significância de cada comparação pareada por ser o mais conservador entre aqueles disponibilizados pelo programa de análise.

Estratégias específicas adotadas na análise de cada um dos parâmetros acústicos estudados são reportadas na seção dedicada aos resultados individuais.

3.2 Resultados: Duração

Previamente à aplicação das ANOVA aos dados de duração, eles foram separados em dois grupos. No caso da duração das sílabas, separou-se os dados referentes às sílabas pré-tônicas daqueles referentes à tônica. No caso das unidades V-V, separou-se de um lado as unidades anteriores àquela que contém a vogal da sílaba tônica da palavra-chave e de outro a unidade que contém a referida vogal[3]. Este procedimento foi adotado para evitar que a grande variabilidade observada na posição tônica (que pode receber um alongamento bastante apreciável no caso de ser o ponto culminante do acento frasal) não comprometesse a análise das pré-tônicas, o principal alvo de interesse da presente investigação. Esta decisão foi tomada após a realização de alguns testes estatísticos realizados no *corpus* do estudo-piloto, que mostraram que o alongamento das posições pré-tônicas, tanto considerando as sílabas quanto as unidades V-V, é influenciado pela posição da unidade na cadeia sintagmática do enunciado (fator PS) de

[3]Exemplificando: no caso das palavras com quatro sílabas, *periquito* p. e., analisou-se num grupo as sílabas "pe" e "ri" e em outro a sílaba tônica "qui". No caso das unidades V-V, se a mesma palavra estivesse inserida na frase-veículo "O periquito rude [...]", o grupo das unidades "pré-tônicas" seria composto pelos V-Vs *o# p*, *er* e *iqu*. A unidade V-V "it", contendo a vogal da sílaba tônica "qui", compõe o grupo das unidades "tônicas".

forma mais significativa do que pela posição da unidade considerada relativamente à unidade de maior força no grupo acentual (fator d_σ). Os mesmos testes mostraram, por outro lado, que as posições "tônicas" são diferenciadas de forma bastante significativa pelo fator d_σ, o que justifica o tratamento diferenciado.

Os dados de duração, depois de normalizados, foram agrupados em unidades do tamanho da sílaba, sendo elas a sílaba fonológica e as unidades V-V. Serão apresentados resultados para estas duas unidades. Ao reportar os resultados estatísticos, as seguintes convenções serão utilizadas: as seções iniciadas por (i) reportam os efeitos principais e as interações entre fatores nas ANOVAs. As iniciadas por (ii) reportam os resultados dos testes *post hoc* e o uso do símbolo "$\neq$" quer dizer que a diferença entre os 2 ou mais termos comparados é significativa ($\alpha = 5\%$).

3.2.1 Influência dos fatores PS e d_σ

Os resultados de duração das sílabas e das unidades V-V serão apresentados em quatro grupos, de acordo com os valores da variável independente PS. As palavras-chave que formam cada grupo e o número de sílabas e unidades V-V analisadas são mostradas na tabela 3.2.

grupo	sílabas	unidades V-V	palavras
A	4	5	*arataca* *patarata* *iridite* *periquito* *urucaca* *pururuca* *horroroso* *gonococo* *edesseno*
B	5	6	*arapabaca* *jaratacaca* *pirimidina* *urucubaca* *hodofobia* *locomotiva*
C	5	6	*colonoscopia*[4]
D	7	8	*dirigibilidade*

Tabela 3.2: Número de sílabas e unidades V-V analisadas em cada grupo de palavras-chave no experimento de produção I.

[4] O sujeito realizou como ditongo o encontro vocálico "ia" na palavra *colonoscopia*. Como as formas ditongadas foram medidas como um só intervalo acústico, a contagem de posições resultou em 5 sílabas, /ko/$_1$ /lo/$_2$ /nos/$_3$

Os resultados mais robustos são os relativos aos grupos A e B, pois nestes grupos há mais variedade de contextos segmentais e consequentemente maior número de dados. Sendo assim, as médias destes grupos tendem a neutralizar as diferenças de duração devidas à duração intrínseca dos segmentos, podendo então os valores médios de cada unidade silábica ou V-V, bem como o contorno que a sucessão destas diversas unidades definem, ser considerados como uma aproximação daquilo que o modelo dinâmico do ritmo pretende simular — "a kind of guiding grid for the superposing segmental gestures" (Barbosa 2002b). Os resultados dos grupos C e D, compostos de apenas uma palavra cada um, devem ser considerados com ressalvas, pois as médias em cada um podem refletir a especificidade da composição segmental da palavra que o compõe e não a organização temporal mais geral do grupo acentual.

Grupo A

Sílabas (i) fator PS: efeito significativo ($F(1,870) = 99,9\ p < 10^{-4}$); fator d_σ e interação: não significativos; (ii) posição 1 $\neq$ da 2, para os níveis 3 e 4 do fator d_σ. Ver figura 3.1a na página 41, mostrando o valor médio de duração normalizada das quatro posições silábicas.

Unidades V-V (i) fator PS: efeito significativo ($F(2,1305) = 456,06\ p < 10^{-4}$); fator d_σ e interação: não significativos; (*ii*) posições 1 e 2 $\neq$ da unidade 3 ($p < 10^{-6}$). Ver figura 3.1b na página 41, mostrando o valor médio de duração normalizada das cinco posições V-V.

Tônica (i) fator d_σ: efeito significativo ($F(4,434) = 14,2\ p < 10^{-5}$); (ii) nível 0 do fator $\neq$ dos demais níveis ($p < 10^{-2}$).

Unidade V-V$_4$ (i) fator d_σ: efeito significativo ($F(4,434) = 14,2\ p < 10^{-5}$); (ii) nível 0 do fator $\neq$ dos demais níveis ($p < 10^{-2}$, no mínimo).

Grupo B

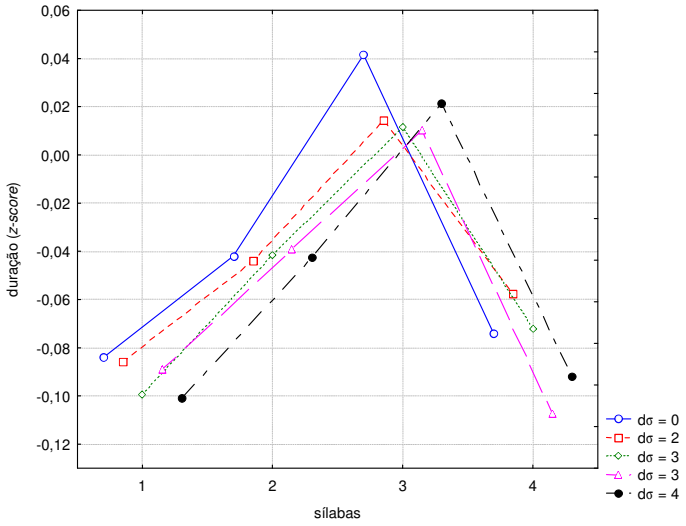
Sílabas (i) fator PS: efeito significativo ($F(2,858) = 69,26\ p < 10^{-4}$), fator d_σ e interação: não significativos; (ii) posição 1 $\neq$ da 3 para níveis 2, 3 e 4 do fator d_σ. Ver figura 3.2a na página 42, mostrando o valor médio de duração normalizada das cinco posições silábicas.

Unidades V-V (i) fator PS: efeito significativo ($F(3,1144) = 124,3\ p < 10^{-4}$), fator d_σ e interação: não significativos; (ii) posições 2 e 3 são $\neq$ da 4 para todos os níveis do fator d_σ. Ver figura 3.2b na página 42, mostrando o valor médio de duração normalizada das seis posições V-V.

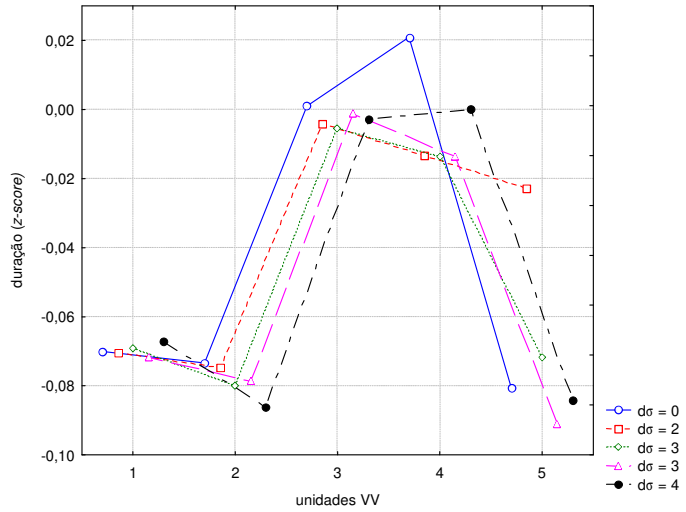
Tônica (i) fator d_σ: efeito significativo ($F(4,286) = 4,7\ p < 10^{-2}$); (ii) nível 0 do fator $\neq$ do 3 ($p < 10^{-2}$).

Unidade V-V$_5$ (i) fator d_σ: efeito significativo ($F(4,544) = 5,46\ p < 10^{-3}$); (ii) nível 0 do fator $\neq$ dos níveis 2 e 3 ($p < 10^{-2}$, pelo menos).

/co/$_4$ /pia/$_5$, e 6 unidades V-V/a# k/$_1$ /ol/$_2$ /on/$_3$ /osk/$_4$ /op/$_5$ /ia# p/$_6$.

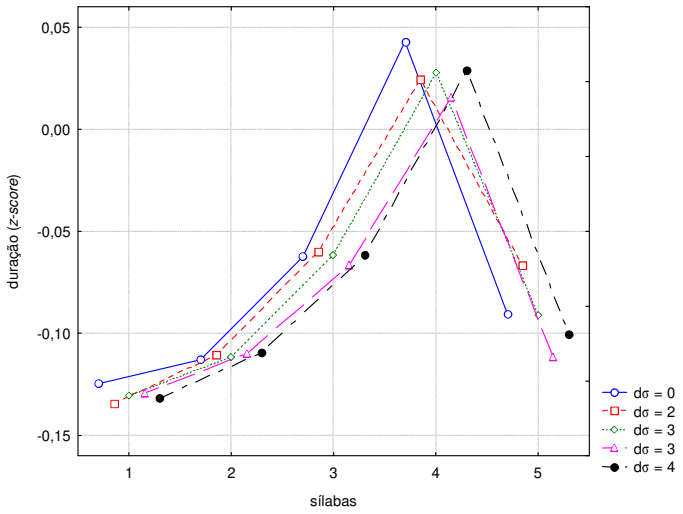


(a) Contorno duracional médio (*z-score*) das sílabas das palavras com 4 sílabas.

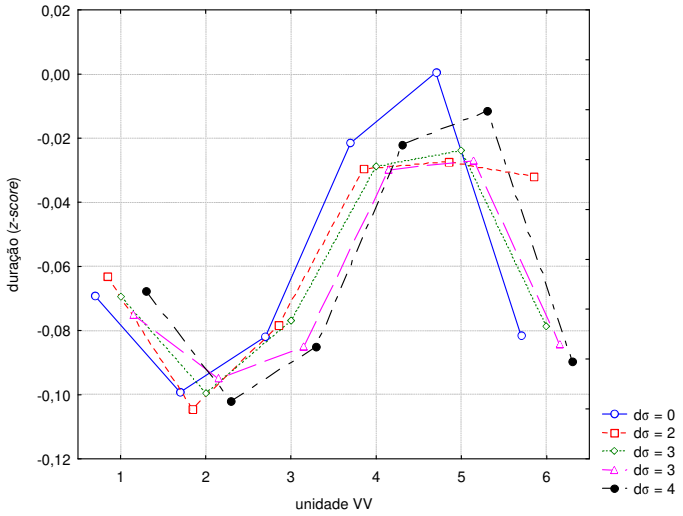


(b) Contorno duracional médio (*z-score*) das unidades V-V das palavras com 5 unidades V-V.

Figura 3.1



(a) Contorno duracional médio (*z-score*) das sílabas das palavras com 5 sílabas.



(b) Contorno duracional médio (*z-score*) das unidades V-V das palavras com 6 unidades V-V.

Figura 3.2

Grupo C

O contorno duracional médio das sílabas e unidades V-V é mostrado nas figuras 3.3a e 3.3b.

Sílabas (i) fator PS: efeito significativo ($F(3,176) = 150,7\ p < 10^{-4}$), fator d_σ e interação: não significativos; (ii) para os níveis 0 e 4 do fator d_σ, a posição 1 $\neq$ da 4 ($p < 10^{-2}$, pelo menos); para os níveis 2 e 3, a posição 1 $\neq$ das posições 2 e 4 e sílaba 3 $\neq$ da 4.

Unidades V-V (i) fator PS: efeito significativo ($F(4,220) = 82,7\ p < 10^{-4}$), fator d_σ não significativo; interação: marginalmente significativa ($F(16,220) = 1,67\ p = 0,052$); (ii) para todos os níveis do fator d_σ, a posição 1 é $\neq$ das demais ($p < 10^{-2}$, pelo menos); para níveis 0 e 4 do fator d_σ, a posição 2 $\neq$ da 5 ($p < 10^{-3}$).

Tônica (i) fator d_σ: efeito significativo ($F(4,44) = 3,06\ p = 0,026$); (ii) nível 0 do fator $\neq$ do 4($p < 10^{-2}$).

Unidade V-V$_5$ (i) fator d_σ: efeito significativo ($F(4,44) = 6\ p < 10^{-3}$); (ii) níveis 0 e 2 do fator $\neq$ do nível 4.

Grupo D

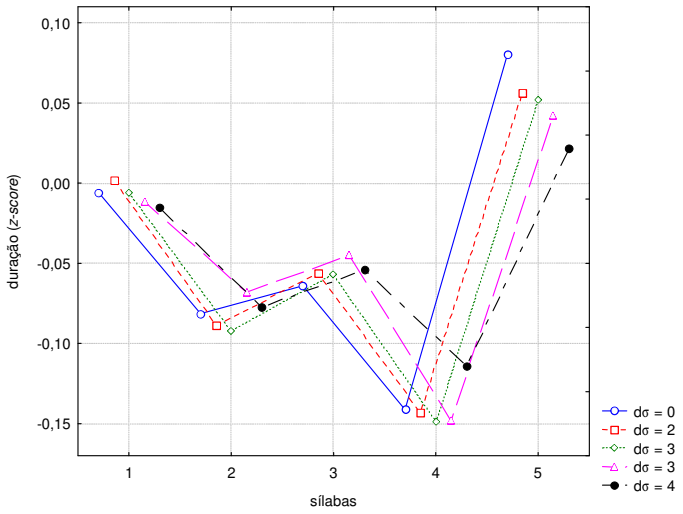
O contorno duracional das sílabas e unidades V-V é mostrado nas figuras 3.4a e 3.4b. Nas duas figuras deve-se observar que as sílabas tônicas apresentam valores inesperadamente altos. Isto se deve a uma falha no processo de normalização dos dados de duração que será explicada a seguir. O locutor cujos dados de duração serviram para estimar os valores de referência usados no cálculo do *z-score* não realiza a africação de [t d] diante de [i] que ocorre sistematicamente na fala do sujeito do experimento. Como as versões africadas das consoantes oclusivas ([t͡ʃ] e [d͡ʒ]) tendem a apresentar durações maiores do que as realizações não-africadas ([t] e [d]), o valor de *z-score* nestes casos será artificialmente alto.

Sílabas (i) fator PS: efeito significativo ($F(4,220) = 151,8\ p < 10^{-4}$); fator d_σ e interação: não significativos; (ii) sílaba 1 $\neq$ das sílabas 2, 3 e 4 para todos os níveis de d_σ e para o nível 0, a $\neq$ extende-se à todas as sílabas ($p = 0,02$, no mínimo); para o nível 2 de d_σ a sílaba 3 $\neq$ 5 ($p < 10^{-3}$). Ver figura 3.4a na página 45, mostrando o valor médio de duração normalizada das sete posições silábicas.

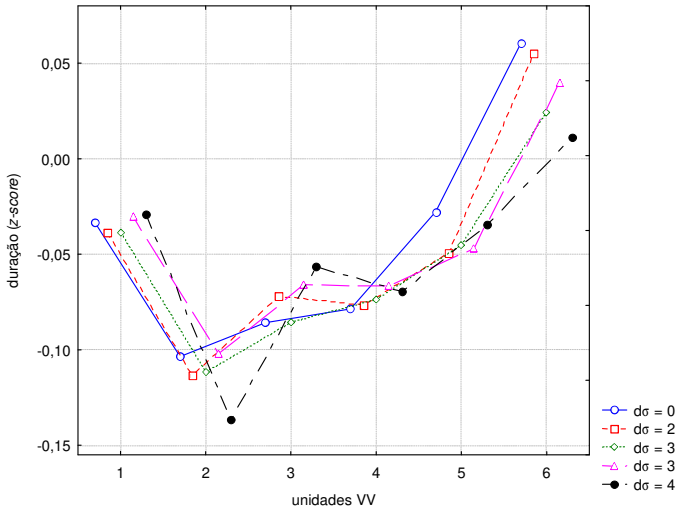
Unidades V-V (i) fator PS: efeito significativo ($F(5,264) = 65,6\ p < 10^{-4}$); fator d_σe interação: não significativos. (ii) para o nível 3 (oxítona) do fator d_σ as unidades 1 e 3 são diferentes significativamente ($p = 0,028$). Ver figura 3.4b na página 45, mostrando o valor médio de duração normalizada das oito posições V-V.

Tônica (i) fator d_σ: não significativo; (ii) nenhuma $\neq$.

Unidade V-V$_7$ (i) fator d_σ: efeito significativo ($F(4,44) = 4,62\ p < 10^{-3}$); (ii) níveis 3 oxítona ($p = 0,02$) e paroxítona ($p = 0,01$) $\neq$ do nível 0.

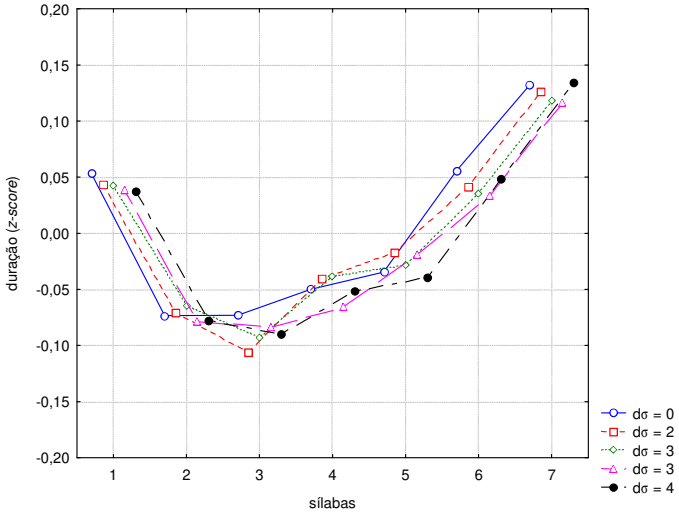


(a) Contorno duracional médio (*z-score*) das sílabas da palavra *colonoscopia*.

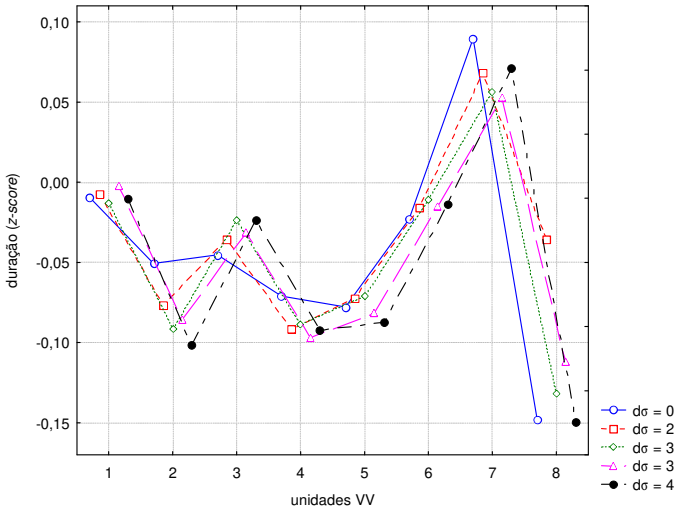


(b) Contorno duracional médio (*z-score*) das unidades V-V da palavra *colonoscopia*.

Figura 3.3



(a) Contorno duracional médio (*z-score*) das sílabas da palavra *dirigibilidade*.



(b) Contorno duracional médio (*z-score*) das unidades V-V da palavra *dirigibilidade*.

Figura 3.4

3.2.2 Influência da natureza das vogais (VR)

No grupo A, composto pelas 9 palavras com 4 sílabas, procedeu-se a uma ANOVA de uma entrada nos dados não normalizados de duração da rima, tendo VR como fator. Limitou-se os dados àqueles correspondentes às posições pré-tônicas, pois somente nestas posições foi possível manter o controle estrito de todas as vogais iguais na rima silábica.

Como se poderia esperar, o fator mostra-se significativo ($F(4,868) = 36,258$ $p < 10^{-4}$), mostrando que efetivamente há uma influência da vogal na duração da rima. Esta influência presente no *corpus* está de acordo com aquilo que se pode considerar como tendência universal na manifestação da duração intrínseca (Lehiste 1970). A vogal baixa /a/ apresenta a maior média e as vogais altas /u i/, valores menores relativamente à vogal baixa. No *corpus*, a seguinte ordenação dos valores das médias de duração aparece: a > o > i = u > e. Os testes de comparação *post hoc* mostram que todas as diferenças entre médias são significativas (pelo menos $p = 0,02$) a menos da comparação entre /i/ e /u/.

No grupo B, composto pelas 6 palavras com 5 sílabas, procedeu-se a uma ANOVA de uma entrada nos dados não normalizados de duração da rima nas posições correspondentes às sílabas pré-tônicas, tendo VR como fator. Neste grupo, assim como no grupo A, somente nas posições correspondentes às pré-tônicas foi possível ter-se sempre a mesma vogal nas rimas.

Neste grupo, em nova semelhança com o outro, o fator resultou significativo ($F(3,861) = 157,94$ $p < 10^{-4}$). O ordenação entre as médias é a seguinte: a > i = o = u. As comparações *post hoc* mostram ser a vogal /a/ distinta de todas as demais vogais ($p = 0,02$, no pior dos casos), /i/ $\neq$ /u/ ($p = 0,03$). Não há diferença entre /o/ e as duas vogais altas.

Estes resultados que mostram a influência das características específicas das vogais na duração da rima, já esperada pelo seu caráter universal, devem ser complementados por outros testes, pois em termos de assinalar eventos acentuais importantes, a duração da rima, sozinha, não constitui contribuição suficiente. É necessário, portanto, verificar se o fator VR tem uma influência significativa tanto no contorno duracional das sílabas quanto no das unidades V-V.

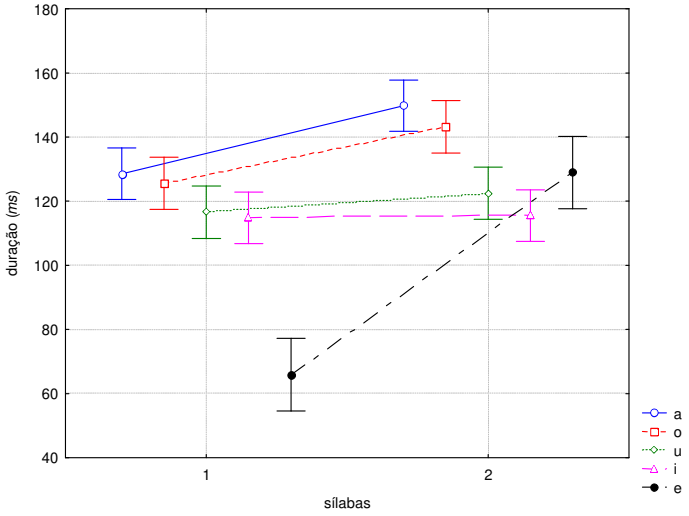
Assim, no que vem sendo chamado de grupo A, efetuou-se uma ANOVA de três entradas nos dados de duração normalizada e não normalizada das sílabas e das unidades V-V[5], tendo como fatores PS, d_σ e VR. Pelos mesmos motivos expostos anteriormente, limitou-se os dados àqueles correspondentes às posições pré-tônicas (posições 1 e 2 no caso da sílaba e posições 2 e 3 no caso das unidades V-V[6]).

No caso das sílabas, o fator VR resulta significativo nos dois tipos de duração ($F_{ms}(4,830) = 23,583$ $p < 10^{-4}$) e ($F_z(4,830) = 41,815$ $p < 10^{-4}$). Mesmo a interação entre os fatores VR e PS resulta também significativa ($F_{ms}(4,830) = 11,027$ $p < 10^{-5}$) e ($F_z(4,830) = 16,207$ $p < 10^{-5}$). No entanto, a interação entre os três fatores mostra-se não significativa para os dois tipos de duração ($F_{ms}(16,830) = 0,082$ ns) e ($F_z(16,830) = 0,28$ ns).

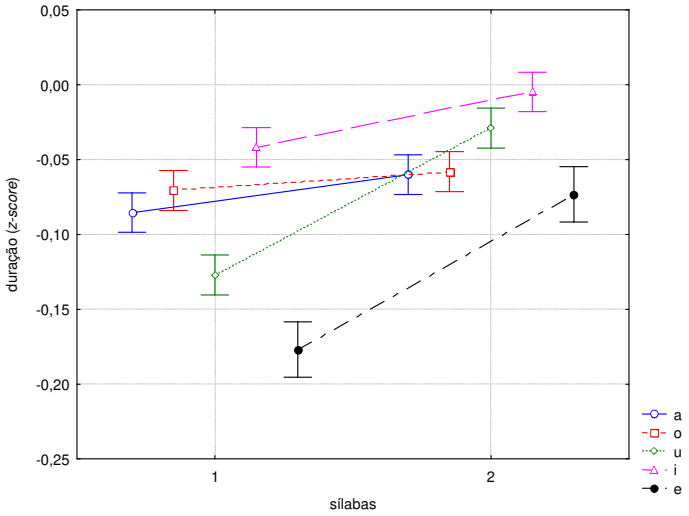
No caso das unidades V-V, o fator VR é significativo nos dois tipos de duração ($F_{ms}(4,830) = 71,090$ $p < 10^{-4}$) e ($F_z(4,870) = 135,40$ $p < 10^{-4}$). A interação entre os fatores VR e PS resulta

[5]Nos casos em que houver os dois tipos de duração, quando a estatística F for reportada, os símbolos subscritos z e ms indicarão se a estatística se refere, respectivamente, aos dados normalizados ou aos não normalizados.

[6]A posição 1, no caso das unidades V-V, corresponde sempre à vogal do artigo masculino ou feminino (vogais /o/ e /a/, respectivamente) que precede as palavras-chave nas sentenças-veículo.

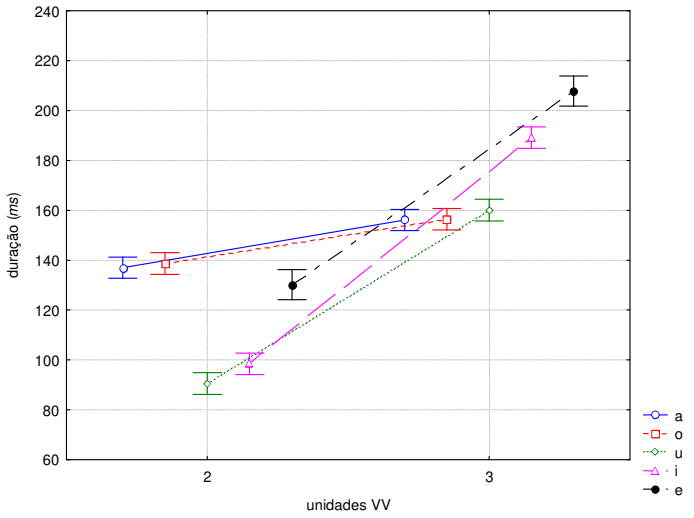


(a) Contornos duracionais (*ms*) para cinco vogais nas duas primeiras sílabas, no conjunto de palavras de 4 sílabas.

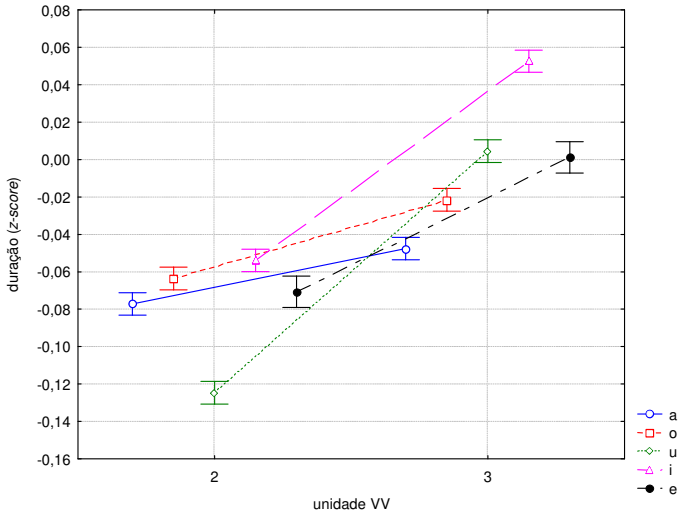


(b) Contornos duracionais normalizados para cinco vogais nas duas primeiras sílabas, no conjunto de palavras de 4 sílabas.

Figura 3.5



(a) Contornos duracionais (*ms*) para cinco vogais em duas unidades V-V, no conjunto de palavras de 4 sílabas.



(b) Contornos duracionais normalizados para cinco vogais em duas unidades V-V, no conjunto de palavras de 4 sílabas.

Figura 3.6

significativa ($F_{ms}(4,830) = 114,23\ p < 10^{-4}$) e ($F_z(4,830) = 94,008\ p < 10^{-4}$). Mais uma vez, a interação entre os três fatores mostra-se não significativa para os dois tipos de duração ($F_{ms}(16,830) = 0,4\ ns$) e ($F_z(16,830) = 1,08\ ns$).

No grupo *B*, efetuou-se uma ANOVA de três entradas nos dados de duração normalizada e não normalizada das sílabas e das unidades V-V, tendo como fatores PS, d_σ e VR. Limitou-se os dados àqueles correspondentes às posições pré-tônicas (posições 1 a 3 no caso da sílaba e posições 2 a 4 no caso das unidades V-V.

Nas sílabas, o fator VR resulta significativo para os dois tipos de duração ($F_{ms}(3,813) = 50,645\ p < 10^{-4}$) e ($F_z(3,813) = 109,76\ p < 10^{-4}$). A interação de VR e PS é também significativa nos dois casos de duração ($F_{ms}(6,813) = 131,70\ p < 10^{-4}$) e ($F_z(6,813) = 116,57\ p < 10^{-4}$). A interação entre os três fatores não é significativa para as duas modalidades de duração ($F_{ms}(24,813) = 0,183\ ns$) e ($F_z(24,813) = 0,406\ ns$).

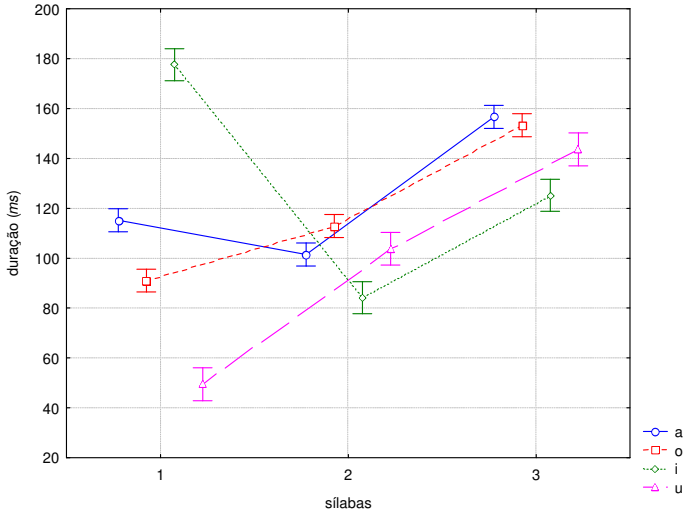
Nas unidades V-V, o fator VR alcançou significância nas duas modalidades de duração ($F_{ms}(3,813) = 57,782\ p < 10{-}4$) e ($F_z(3,813) = 64,87\ p < 10{-}4$). A interação VR e PS também mostrou-se significante ($F_{ms}(6,813) = 131,43\ p < 10^{-4}$) e ($F_z(6,813) = 125,52\ p < 10^{-4}$. A interação entre os três fatores é significativa ($F_{ms}(24,813) = 0,44\ ns$) e ($F_z(24,813) = 0,72\ ns$).

Pode-se interpretar este conjunto de resultados conforme o seguinte princípio (que leva em conta uma relação hierárquica entre níveis de duração segmental e prosódica): apesar do fenômeno de duração intrínseca das vogais continuar a exibir sua influência, como mostram os resultados no caso do fator VR e da interação entre este fator e PS, pode-se observar nas figuras 3.5a–3.8b, nas páginas 51–47, que esta influência não está em desacordo com a tendência geral de aumento de duração que se observa da primeira posição para a segunda tanto no caso da sílaba quanto no caso de V-V. Prova disso é que quando se acrescenta o fator d_σ na interação, ela resulta não significativa, mostrando a coerência entre os contornos nas diversas condições de acento frasal.

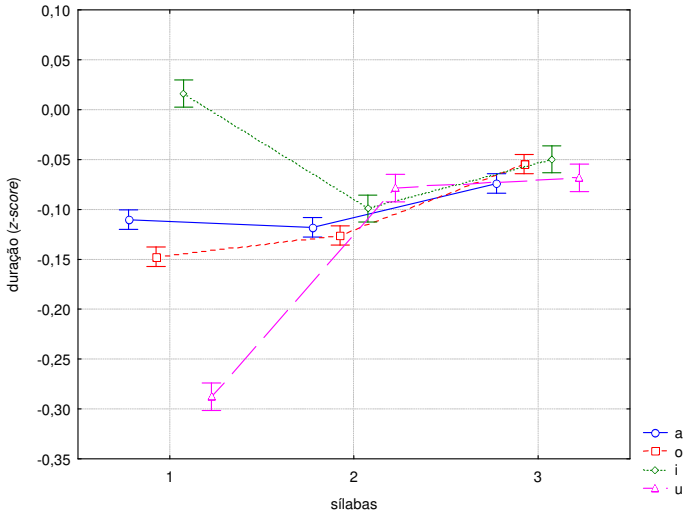
O que este controle permite discutir a respeito do modelo do ritmo é que, apesar de influências como a duração intrínseca dos segmentos, o contorno duracional do grupo acentual pode ser caracterizado em função da sucessão de mudanças no grau de alongamento (medido em *z-scores*) associado a cada unidade V-V. A cada posição na cadeia sucessiva de unidades V-V, que pode ser pensada nos termos do modelo dinâmico como cada ciclo do oscilador silábico induzido, associa-se um grau de alongamento dos gestos ou conjunto de gestos articulatórios que são mobilizados para a realização daquele ciclo. O grau de alongamento sempre dependerá da proximidade do ciclo em questão em relação à batida do oscilador frasal e também por fatores como a natureza dos gestos a ser realizado. No entanto, a forma qualitativa do contorno duracional dos grupos acentuais mostra ser menos afetada por perturbações como esta e mais definida pela sucessão de aumentos e diminuições do alongamento dos ciclos V-V ou nos termos do modelo, do período do oscilador silábico induzido.

3.2.3 Influência da sílaba inicial ES

A inclusão da variável ES deve servir, como foi explicado na seção 3.1.1, para testar uma intuição de Mattoso Câmara Jr. sobre as sílabas sem ataque consonantal serem menos propensas a receber proeminências. Para esclarecer a questão é necessário contrastar duas situações. No

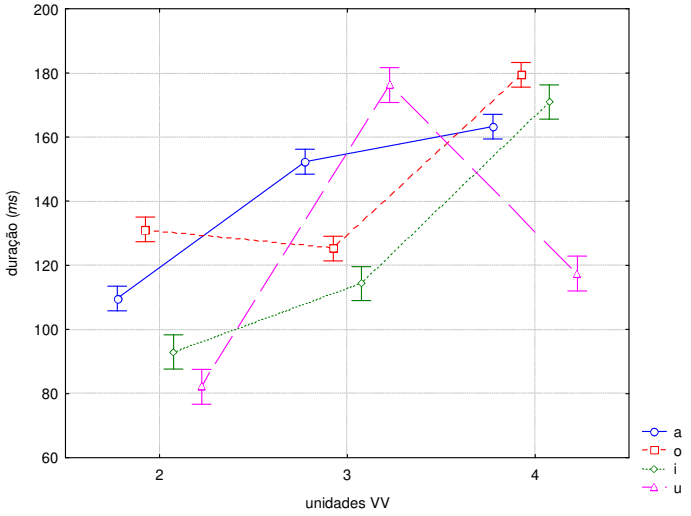


(a) Contornos duracionais (*ms*) para quatro vogais nas três primeiras sílabas, no conjunto de palavras de 5 sílabas.

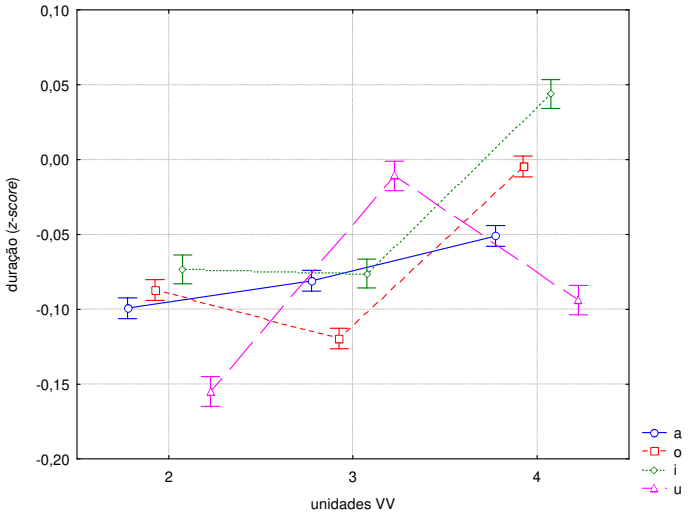


(b) Contornos duracionais normalizados para quatro vogais nas três primeiras sílabas, no conjunto de palavras de 5 sílabas.

Figura 3.7



(a) Contornos duracionais (*ms*) para quatro vogais em três três unidades V-V, no conjunto de palavras de 5 sílabas.



(b) Contornos duracionais normalizados para quatro vogais em três três unidades V-V, no conjunto de 6 palavras de 5 sílabas.

Figura 3.8

corpus, estas situações podem ser exemplificadas pelo par "As aratacas" e "A patarata". O trecho do grupo acentual que interessa contrastar são as duas primeiras unidades V-V de cada um, ou seja, /as/$_1$ /ar/$_2$ *versus* /ap/$_1$ /at/$_2$. Comparando os dois contornos duracionais será possível verificar a influência da estrutura da sílaba inicial (sílaba iniciada por consoante *versus* sílaba iniciada por vogal) na forma do contorno duracional do grupo acentual.

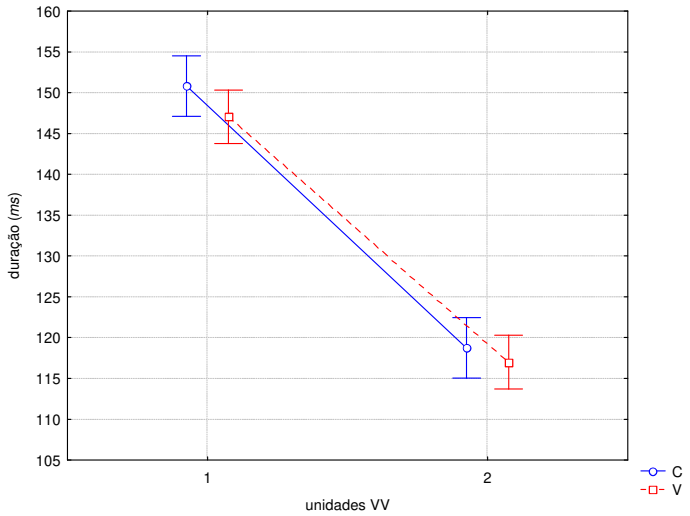
Realizou-se, nos dados de duração normalizados e não normalizados, uma ANOVA de três entradas tendo ES, PS e d_σ como fatores. Pelas razões expostas no parágrafo anterior, restringiu-se o fator PS às posições 1 e 2. Far-se-á referência aos dois níveis do fator ES, sílaba iniciada por consoante ou sílaba iniciada por vogal, como condição C ou V, respectivamente.

No grupo A, o fator ES não resultou significativo para as duas modalidades de duração – ($F_{ms}(1,860) = 2,36\ ns$) e ($F_z(1,860) = 0,5\ ns$). A interação entre o fator ES e PS não resultou significativa para os dados brutos ($F_{ms}(1,860) = 0,33\ ns$), mas sim no caso dos dados normalizados ($F_z(1,860) = 7,22\ p < 10^{-2}$). No caso dos dados brutos, as comparações *post hoc* mostram que as condições C e V não são significativamente diferentes se tomadas posição a posição. Somente resultam significativas as comparações entre diferentes posições ($p < 10^{-6}$, em todos os casos), independente da condição de ES. No caso dos dados normalizados, as condições C e V também não são significativamente diferentes se tomadas posição a posição. No entanto, ao contrário da condição C, na condição V a diferença entre as posições não é significativa. A interação entre os três fatores não é significativa – ($F_{ms}(4,860) = 0,12\ ns$) e ($F_z(4,860) =, 045\ ns$). Ver figura 3.9a e 3.9b, que mostram os valores médios de duração bruta e normalizada das duas primeiras posições V-Vno grupo A para os dois níveis do fator ES.

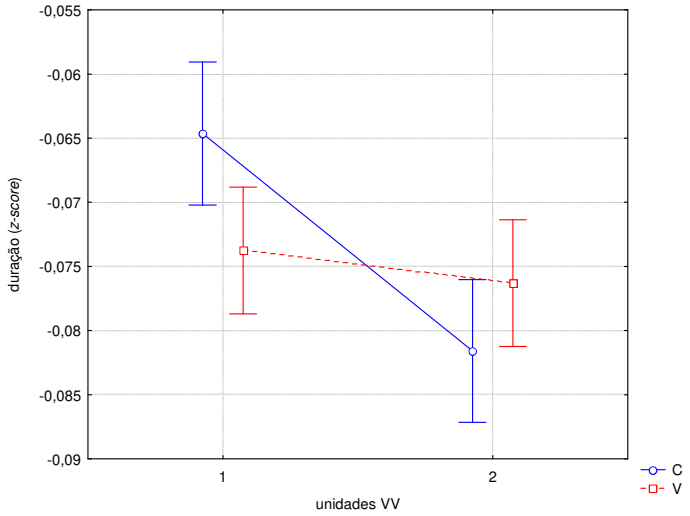
No grupo B, o fator ES não resultou significativo para os dados brutos ($F_{ms}(1,562) = 0,112\ ns$), mas alcançou a significância para os normalizados ($F_z(1,562) = 14,75\ p < 10{-}3$). A interação entre o fator ES e PS é significativa para as duas modalidades de duração – ($F_{ms}(1,562) = 31,4\ p < 10^{-6}$) e ($F_z(1,562) = 52,13\ p < 10^{-5}$. No caso dos dados brutos todas as comparações *post hoc* são significativas ($p < 10^{-2}$, no mínimo). No caso dos dados normalizados, para as duas posições os valores da condição C não são significativamente diferentes. O contrário ocorre com a condição V, em que os valores para as duas posições é significativamente diferente ($p < 10^{-6}$). Na posição 1, as condições C e V não são significativamente diferentes, ao contrário do que ocorre na posição 2 ($p < 10^{-6}$). A interação entre os três fatores não é significativa em nenhum modo de duração – ($F_{ms}(4,562) = 0,72\ ns$) e ($F_z(4,562) = 0,9\ ns$). Ver figuras 3.10a e 3.10b, que mostram os valores médios de duração bruta e normalizada das duas primeiras posições V-V no grupo B para os dois níveis do fator ES.

Estes resultados mostram que a influência do fator ES não se manifesta de maneira muito regular nos dois grupos analisados e nas duas modalidades de duração analisadas.

No caso do grupo A, há uma divergência entre o comportamento se comparados os resultados para as duas modalidades de duração. Nos dados brutos tem-se um comportamento idêntico para ambas as condições de ES (C e V) nas duas posições testadas. Observa-se uma queda na duração da primeira para a segunda, mas as duas condições de ES acompanham indistinguivelmente a queda. Já no caso dos dados normalizados, embora a condição ES não seja uma influência significativa, na interação com o fator PS temos uma complicação da situação, pois embora na primeira posição o valor de cada condição de ES seja indistinguível, na segunda somente observa-se uma queda estatisticamente significativa no valor da condição C.

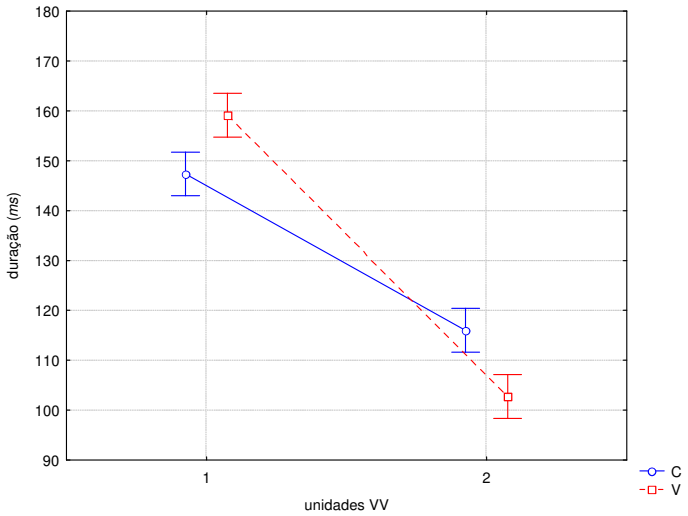


(a) Contornos duracionais (*ms*) para as condições C e V nas duas primeiras unidades V-V, no conjunto de palavras de 4 sílabas.

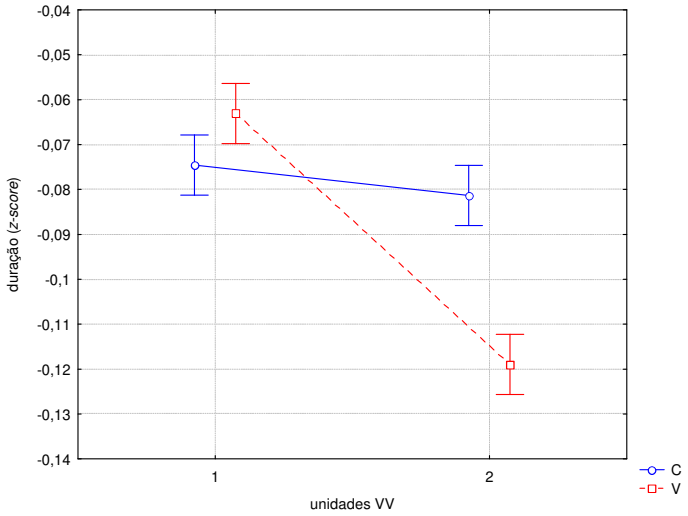


(b) Contornos duracionais normalizados para as condições C e V nas duas primeiras unidades V-V, no conjunto de palavras de 4 sílabas.

Figura 3.9



(a) Contornos duracionais (*ms*) para as condições C e V nas duas primeiras unidades V-V, no conjunto de palavras de 5 sílabas.



(b) Contornos duracionais normalizados para as condições C e V nas duas primeiras unidades V-V, no conjunto de 6 palavras de 5 sílabas.

Figura 3.10

Sendo esta a situação, pode-se pensar que, embora a posição 2, que contém a primeira vogal da palavra-chave, não seja estatisticamente diferente sob as duas condições do fator ES, é possível que se consiga um efeito de proeminência relativa em favor da primeira unidade V-V da palavra seguinte através de um alongamento diferenciado da posição 1, no caso da condição C. Neste caso, o primeiro ciclo V-V, que contém a vogal do artigo mais a primeira consoante da palavra receberia um alongamento maior do que seria o caso na condição V, que contém sempre o artigo mais o morfema *-s* do plural.

No caso do grupo *B*, no entanto, não se observa uma repetição sistemática do padrão encontrado para o grupo *A*. Somente os dados não normalizados mostram uma situação parcialmente semelhante, pois da primeira para a segunda posição há uma queda no valor de ambas as condições de ES. Mesmo havendo esta coerência, a comparação entre as condições C e V em cada posição separadamente resulta significativa. O desacordo maior observa-se nos dados normalizados, para os quais a situação em relação ao grupo *A* inverte-se. Na posição 1, como no caso do grupo *A*, C e V não apresentam diferença significativa. Já na posição 2, ao contrário do grupo *A*, a condição V apresenta uma queda em relação a C que é significativa. Também em contradição com o resultado do grupo *A*, é a condição V que não apresenta diferença estatística entre os valores para a posição 1 e 2. Assim, para o grupo *B* não pode valer a mesma análise feita para o grupo *A*. No grupo de palavras de 5 sílabas (*B*), também é possível ver a formação de uma proeminência relativa em favor da posição 1, só que diferentemente do grupo *A*, ela se dá em função da diminuição da duração na posição 2 e para a condição V. A consequência disto é que, na posição 2, o fator ES atua de forma significativa, mas ao contrário do que seria esperado, uma vez que a proeminência se dá em favor das palavras sem ataque consonantal na primeira posição. A regularidade observada é apenas a existência de uma proeminência inicial, realizada segundo estratégias diferentes. Ora por uma duração maior do ciclo V-V inicial, ora por uma duração menor do segundo ciclo V-V da sequência.

Em termos da discussão do modelo de ritmo, o que este controle permitiu mostrar é que o alongamento da unidades V-V não se mostra sensível à afiliação silábica da consoante que segue a vogal na unidade V-V, o que conta como evidência para o caráter não lexical das relações prosódicas. Em outras palavras, o contorno de duração imposto à cadeia fônica ignora fronteiras lexicais.

3.2.4 Discussão parcial

Os resultados referentes ao parâmetro duração relatados nesta seção representam uma contribuição para a elucidação da natureza do acento secundário no PB, ao mesmo tempo em que ajudam a entender a participação de certos parâmetros acústicos no processo de acentuação.

Especificamente quanto ao acento secundário, uma das questões que os resultados ajudam a verificar é a previsão empírica, derivada das análises fundadas nas teorias métricas, de uma alternância binária entre as sílabas pré-tônicas. Para o parâmetro acústico duração, não se encontra, nos resultados dos experimentos aqui relatados, sustentação para a afirmação de uma tal alternância sinalizada por aquele parâmetro. Os padrões de mudança da duração ao longo da palavra, quer se forem consideradas as sílabas quer as unidades V-V, dificilmente poderiam ser interpretados como a realização de um padrão de alternância binária. Os contornos duracionais das sílabas, considerando os resultados dos grupos *A* e *B*, indicam aumento progressivo da

duração das sílabas até a realização da sílaba tônica, seguindo-se uma queda, correspondente à duração da pós-tônica.

Os padrões de duração das unidades V-V conformam-se àqueles previstos pelo modelo dinâmico do ritmo. Em certos casos, entretanto, foi possível observar a ocorrência de um padrão um pouco diferente do previsto, pois a primeira unidade V-V tem um valor mais alto do que o segundo e só a partir deste observa-se o aumento das unidades V-V seguintes até a culminância frasal. Este padrão em que se observa a primeira unidade V-V proeminente pode ser produzido pelo modelo, como mostram simulações realizadas por Barbosa (2002b). Estas simulações foram precedidas por um ajuste dos parâmetros α e β do modelo, conseguido por meio da aplicação de uma técnica que estima pares de valores para os referidos parâmetros que minimizam as diferenças entre as durações produzidas pelo modelo e os contornos duracionais médios de frases lidas, retiradas de um *corpus* previamente gravado. Posteriormente, procedimentos semelhantes serão adotados para o *corpus* do nosso experimento para que seja confirmado em outra instância o acerto das previsões feitas pelo modelo e das pressuposições teóricas subjacentes à sua proposição.

Os resultados sugerem, nesta discussão, uma influência gradiente da temporalidade extrínseca sobre a temporalidade intrínseca, nos termos da visão dinamicista. A montagem do *corpus* permitiu que esta influência se materializasse, através da variável independente d_σ, na forma de um afastamento gradual da culminância do acento frasal em relação à sílaba tônica da palavra-chave nas frases-veículo. O que os resultados sugerem é que um contínuo físico, isto é, a defasagem gradiente entre a tônica lexical e a culminância do acento frasal, tem como efeito provocar uma alteração, também gradiente, nas relações de fase especificadas entre os gestos na forma lexicalizada da palavra. Como resultado, temos gestos mais ou menos sobrepostos ou temporalmente estendidos, tendo como consequência acústica maior ou menor duração para as unidades do tamanho da sílaba. Como sugestão de que esta suposta influência gradiente é sensível à categoria linguística acento lexical, pode-se citar o fato experimental de que as pré-tônicas, ou ciclos V-V mais distantes da batida do oscilador frasal, são altamente influenciadas pela posição que ocupam na sequência e pouco influenciadas pelo contexto prosódico (variável d_σ). Já as tônicas são bastante afetadas pelo contexto prosódico (sua duração é sempre maior quando coincide com a culminância frasal, tendendo a diminuir quando a culminância ocorre posteriormente). As pós-tônicas também apresentam influência do contexto prosódico, podendo ter valores de duração altos se ocorrerem próximos a uma proeminência frasal (observar, a este respeito, os valores sempre altos do ciclo V-V que contém a vogal pós-tônica quando $d_\sigma = 2$).

3.3 Resultados: Distribuição de acentos frasais

Nesta seção serão reportados os resultados da distribuição dos acentos frasais para os quatro grupos de palavras do *corpus* delimitados a partir da aplicação do algoritmo descrito no apêndice A.

Os resultados de cada grupo de palavras serão apresentados por meio de três gráficos e uma tabela: (a) gráfico de barras que mostra, para cada nível da variável d_σ, o número de acentos frasais por frase-veículo; (b) gráfico de barras que mostra a frequência relativa de ocorrência de acentos frasais em cada unidade V-V da frase-veículo para todos os níveis de d_σ e (c) gráfico

de dispersão que mostra a diferença entre V-V_2 e V-V_1 em função do tamanho do primeiro grupo acentual. A tabela mostra a frequência relativa dos diferentes agrupamentos prosódicos atestados em cada nível da variável d_σ.

3.3.1 Grupo A

Nas palavras do grupo A, a unidade V-V na posição 4 contém a vogal da sílaba lexicalmente acentuada das palavras-chave. As posições 6, 7, 7 e 8, respectivamente, contém a vogal lexicalmente acentuada do adjetivo posposto nos níveis 2, 3o, 3p e 4 da variável d_σ.

0		2		3o		3p		4	
posição	%	posição	%	posição	%	posição	%	posição	%
4	61	1	1	7	7	4	15	1|8	3
1|4	38	4	16	1|7	10	7	2	4|8	62
2|4	1	6	16	2|7	4	1|4	10	1|4|8	35
		1|4	14	4|7	56	1|7	7		
		1|6	7	1|4|7	22	4|7	56		
		4|6	28			1|4|7	22		
		1|4|6	18						

Tabela 3.3: Distribuição de ocorrência relativa de agrupamentos prosódicos nos níveis da variável d_σ na amostra das palavras-alvo do grupo A. Nos agrupamentos, os números indicam a posição da unidade V-V que delimita uma fronteira.

3.3.2 Grupo B

Nas palavras do grupo B, a unidade V-V na posição 5 contém a vogal da sílaba lexicalmente acentuada das palavras-chave. As posições 7, 8, 8 e 9, respectivamente, contém a vogal lexicalmente acentuada do adjetivo posposto nos níveis 2, 3o, 3p e 4 da variável d_σ.

0		2		3o		3p		4	
posição	%	posição	%	posição	%	posição	%	posição	%
5	16	5	4	8	5	5	4	9	2
1|5	77	7	4	1|8	35	1|5	24	1|5	2
2|5	5	1|5	21	5|8	7	1|8	9	1|9	22
1|4|7	2	1|7	41	1|5|8	47	2|8	4	5|9	12
		1|5|7	30	2|5|8	5	5|8	13	1|5|9	60
						1|3|8	2	2|5|9	2
						1|5|8	40		
						2|5|8	4		
						1|3|5|8	2		

Tabela 3.4: Distribuição de ocorrência relativa de agrupamentos prosódicos nos níveis da variável d_σ na amostra das palavras-alvo do grupo B. Nos agrupamentos, os números indicam a posição da unidade V-V que delimita uma fronteira.

3.3.3 Grupo C

Na palavra do grupo C, *colonoscopia*, a unidade V-V na posição 6 contém a vogal da sílaba lexicalmente acentuada da palavra-chave. As posições 7, 8, 8 e 9, respectivamente, contém a vogal lexicalmente acentuada do adjetivo posposto nos níveis 2, 3o, 3p e 4 da variável d_σ.

0		2		3o		3p		4	
posição	%	posição	%	posição	%	posição	%	posição	%
1|6	100	1|7	100	1|8	80	1|6	20	1|9	90
				1|6|8	20	1|8	10	1|3|9	10
						1|3|8	20		
						1|6|8	40		
						1|3|6|8	10		

Tabela 3.5: Distribuição de ocorrência relativa de agrupamentos prosódicos nos níveis da variável d_σ na amostra das palavras-alvo do grupo C. Nos agrupamentos, os números indicam a posição da unidade V-V que delimita uma fronteira.

3.3.4 Grupo D

Na palavra do grupo D, *dirigibilidade*, a unidade V-V na posição 7 contém a vogal da sílaba lexicalmente acentuada da palavra-chave. As posições 9, 10, 10 e 11, respectivamente, contém a vogal lexicalmente acentuada do adjetivo posposto nos níveis 2, 3o, 3p e 4 da variável d_σ.

0		2		3o		3p		4	
posição	%	posição	%	posição	%	posição	%	posição	%
7	33	7	20	7|10	20	7	10	7|11	22
1|7	11	1|7	20	1|7|10	30	1|10	10	1|7|11	22
2|7	11	2|7	40	2|7|10	30	2|7	10	2|7|11	22
3|7	33	3|7	10	3|7|10	20	3|7	10	3|7|11	22
1|3|7	11	1|3|7	10			7|10	10	2|7|9|11	11
						2|7|10	10		

			3|7|10 30	
			1|3|7|10 10	

Tabela 3.6: Distribuição de ocorrência relativa de agrupamentos prosódicos nos níveis da variável d_σ na amostra das palavras-alvo do grupo D. Nos agrupamentos, os números indicam a posição da unidade V-V que delimita uma fronteira.

3.3.5 Discussão parcial

Os dados reportados nesta seção mostram a variabilidade na ocorrência de agrupamentos prosódicos definidos pela ocorrência de acentos frasais. De forma geral, pode-se dizer que a posição inicial (unidade V-V 1) e os acentos lexicais da palavra-chave e do adjetivo posposto na frase-veículo tendem a concentrar a grande maioria dos acentos frasais.

Nos dados do grupo A observa-se que a ocorrência de acento frasal em início absoluto de frase é menor do que nos demais grupos. No grupo B os acentos frasais iniciais ocorrem em pelo menos 80% das unidades V-V de início absoluto (unidade V-V 1) e no grupo C em eles ocorrem 100% dos casos

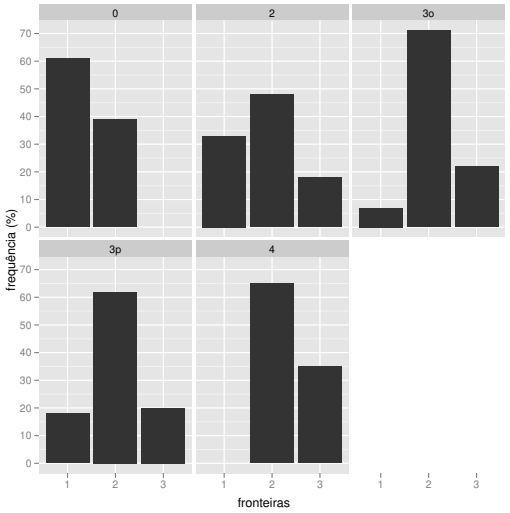
Os grupos C e D são compostos apenas de uma palavra e portanto o menor número de repetições das frases-veículo tendem a tornar os dados mais sujeitos às idiosincrasias da amostra mais restrita. De qualquer forma, os dados da distribuição dos agrupamentos prosódicos apresentados nas tabelas 3.5 e 3.6 mostram que mesmo em cadeias de 4 ou 5 sílabas pré-tônicas a chance de ocorrer um quarto acento frasal (que poderia ser associado a um acento secundário) além de um inicial e daqueles associados às tônicas da palavra-chave e do adjetivo posposto é muito reduzida (1 ocorrências em 50 no grupo C e 2 em 50 no grupo D). Agrupamentos prosódicos em que ocorre um acento frasal na segunda ou terceira posição (representados nas tabelas desta seção como 2|[...] ou 3|[...]) poderiam ser considerados casos de proeminência inicial com anacruse.

3.4 Resultados: Frequência fundamental

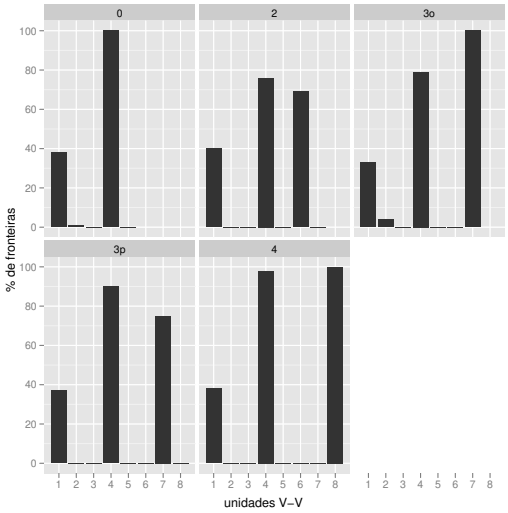
Nas análises estatísticas, a variável dependente é o valor médio da f_0 medida no ponto médio da duração acústica de cada vogal. As vogais da palavra-chave e do adjetivo posposto quando ele existia. Como variáveis independentes haverá os fatores PS, d_σ e VR. O primeiro terá um número de subníveis variável de acordo com o tamanho da palavra-chave. A variável d_σ terá os cinco subníveis listados no exemplo 1 e a variável VR terá cinco subníveis, correspondentes às diferentes vogais que ocupam as rimas nas sílabas pré-tônicas das palavras-chave (cf. tabela 3.1 na página 34).

3.4.1 Análise por valor no ponto médio

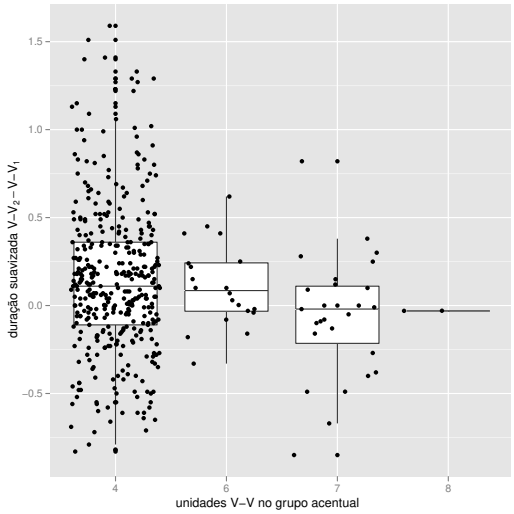
Os resultados são apresentados em quatro grupos, de acordo com os valores da variável independente PS. Os grupos são: A) reúne as 9 palavras com 4 sílabas; B) reúne as 6 palavras



(a) Número de acentos frasais por frase-veículo para todos os valores de d_σ

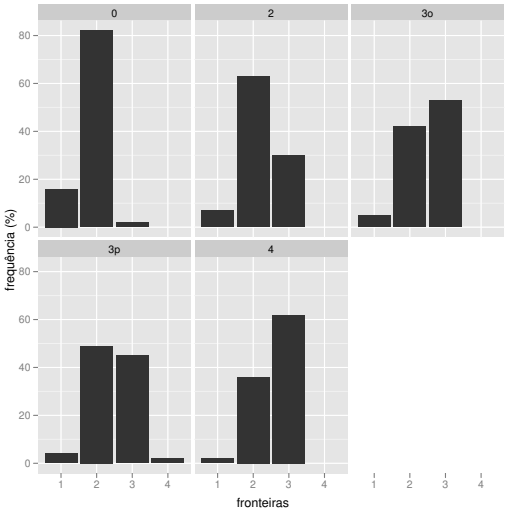


(b) Frequência relativa de ocorrência de acentos frasais em cada unidade V-V da frase-veículo para todos os valores de d_σ

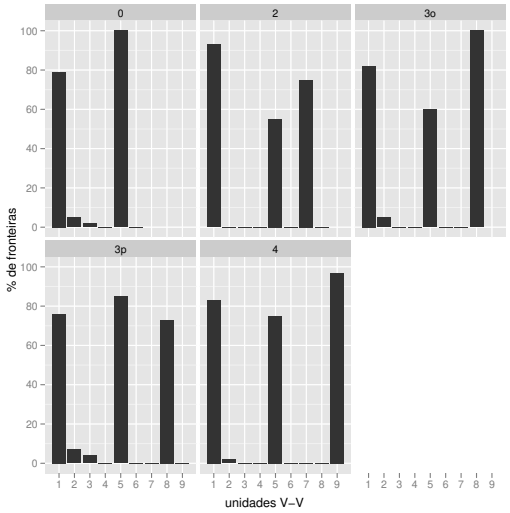


(c) Alongamento inicial em função do tamanho do grupo acentual

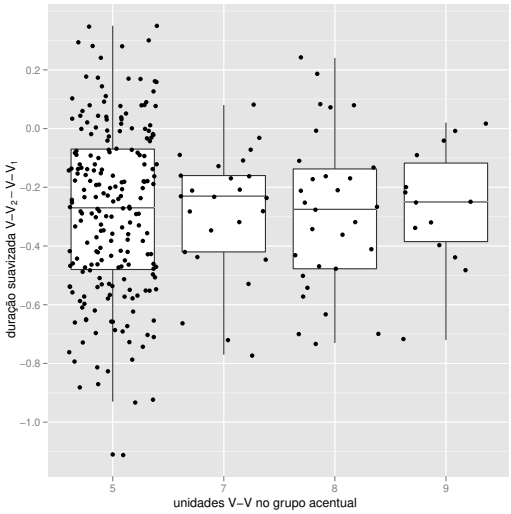
Figura 3.11: Distribuição dos acentos frasais – grupo A



(a) Número de acentos frasais por frase-veículo para todos os valores de d_σ

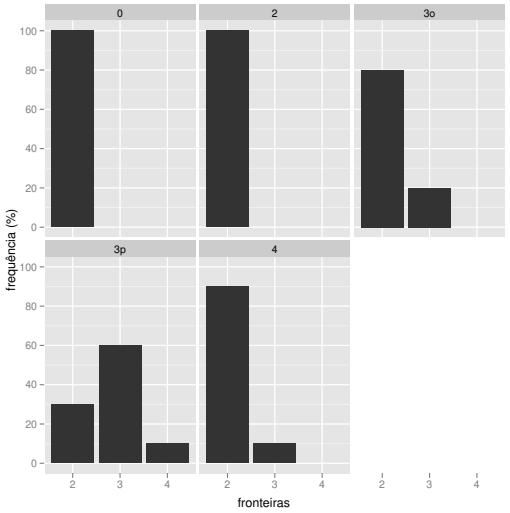


(b) Frequência relativa de ocorrência de acentos frasais em cada unidade V-V da frase-veículo para todos os valores de d_σ

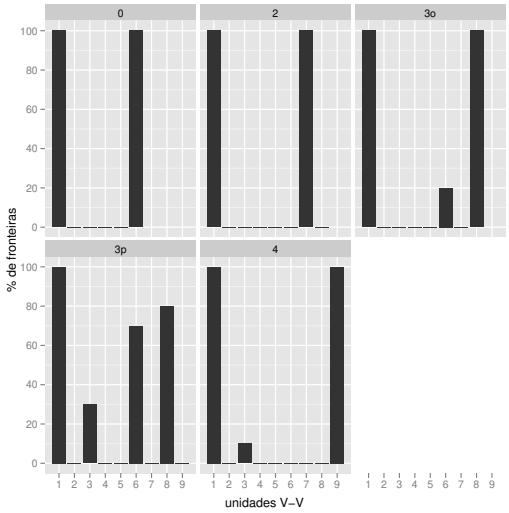


(c) Alongamento inicial em função do tamanho do grupo acentual

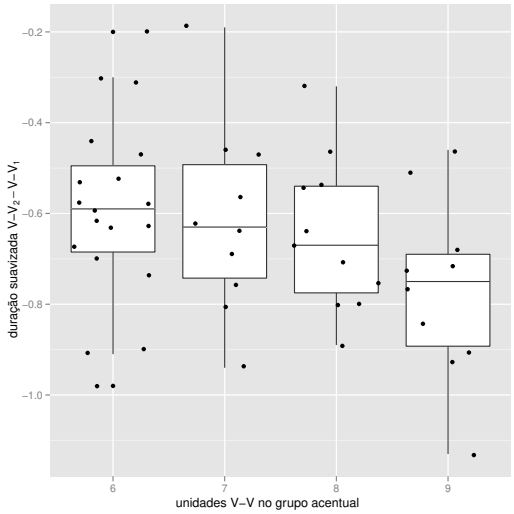
Figura 3.12: Distribuição dos acentos frasais – grupo B



(a) Número de acentos frasais por frase-veículo para todos os valores de d_σ

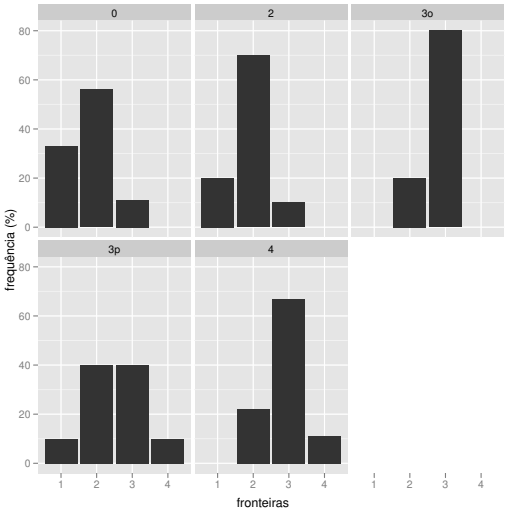


(b) Frequência relativa de ocorrência de acentos frasais em cada unidade V-V da frase-veículo para todos os valores de d_σ

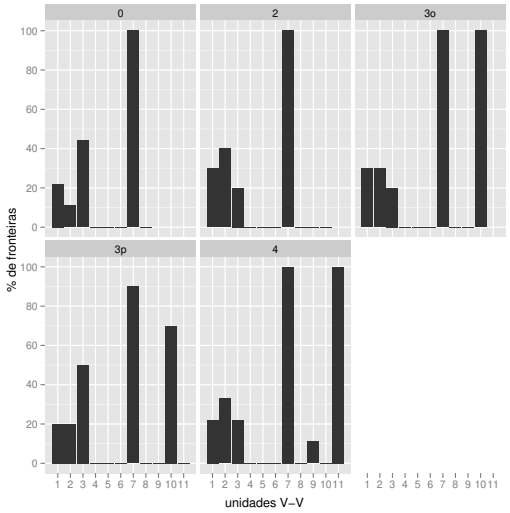


(c) Alongamento inicial em função do tamanho do grupo acentual

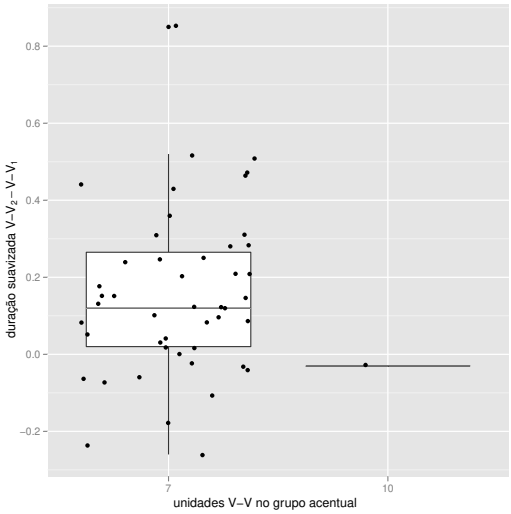
Figura 3.13: Distribuição dos acentos frasais – grupo C



(a) Número de acentos frasais por frase-veículo para todos os valores de d_σ



(b) Frequência relativa de ocorrência de acentos frasais em cada unidade V-V da frase-veículo para todos os valores de d_σ



(c) Alongamento inicial em função do tamanho do grupo acentual

Figura 3.14: Distribuição dos acentos frasais – grupo D

com 5 sílabas; C) palavra *colonoscopia*, com 5 sílabas[7] e D) palavra *dirigibilidade* com 7 sílabas.

Analisaremos em separado o comportamento do parâmetro f_0 nas pré-tônicas para por à prova a hipótese da alternância binária nestas posições.

Grupo A

Pode-se ver na figura 3.15 os valores f_0 no ponto médio das vogais nas palavras do grupo A.

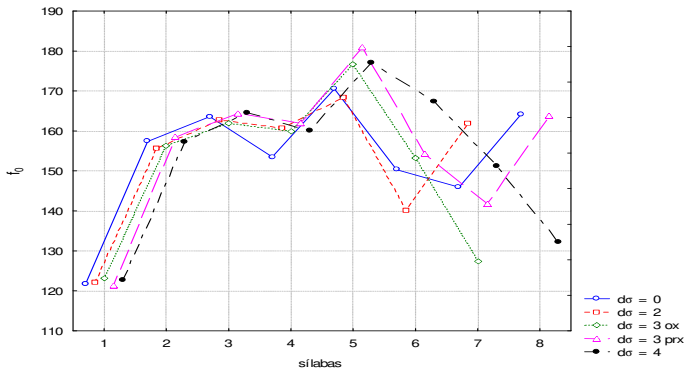


Figura 3.15: Médias de f_0 no ponto médio das vogais nas sentenças-veículo do grupo A. Os padrões de linhas representam os diferentes valores da variável d_σ.

Procedeu-se uma ANOVA de duas entradas nos dados de f_0 relativos ao ponto médio, tendo como fatores PS, restrito às posições 1–4 e d_σ. O efeito devido ao primeiro fator é altamente significativo ($F(2, 1290) = 13,9$ $p < 10^{-6}$) e o efeito resultante do segundo e a interação entre os dois são ambos não significativos — ($F(4, 1290) = 0,9$ ns) e ($F(8, 1290) = 0,9$ ns), respectivamente.

Isto quer dizer que o fator determinante para a definição do contorno nas primeiras posições das sentenças-veículo (as sílabas pré-tônicas e a tônica) é a posição de cada vogal (ou sílaba) na cadeia sintagmática. Testes *post hoc* mostram, no entanto, que no fator PS a diferença entre as posições 2 e 4 não é significativa.

Grupo B

Pode-se ver na figura 3.16 o gráfico de f_0 no ponto médio das vogais nas palavras do grupo.

Nos dados de f_0 no ponto médio uma ANOVAde duas entradas tendo como fatores PS (posições 2–5) e d_σ mostrou que o fator PS produz efeito altamente signinifcativo ($F(3, 1125) = 43,5$ $p < 10^{-6}$), sendo não significativo o efeito do fator d_σ ($F(4, 1125) = 0,8$ ns), bem como a interação entre os dois ($F(12, 1125) = 0,3$ ns). Testes *post hoc* no fator PS mostram que

[7] O sujeito realizou como ditongo o encontro vocálico "ia" na palavra. Como as formas ditongadas foram medidas como um só intervalo acústico, a contagem de posições somou 5 sílabas: co_1 lo_2 nos_3 co_4 pia_5.

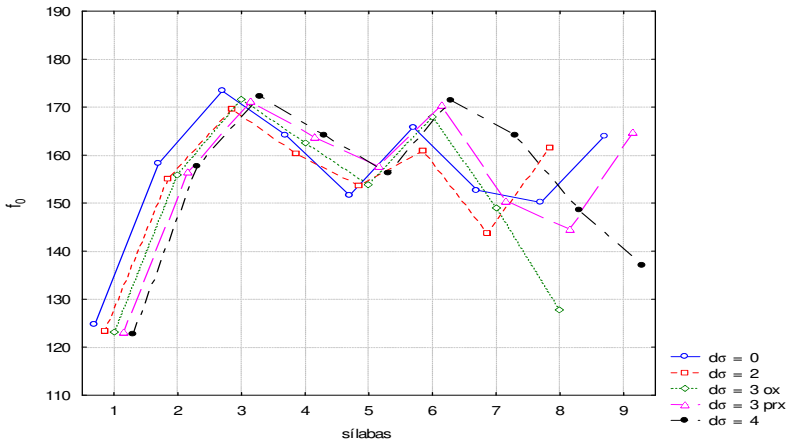


Figura 3.16: Médias de f_0 no ponto médio das vogais nas sentenças-veículo do grupo *B*. Os padrões de linhas representam os diferentes valores da variável d_σ.

a única comparação que não apresenta diferença significativa é aquela entre as posições 2 e 5 (entre a vogal da primeira sílaba pré-tônica e a vogal da sílaba tônica).

Grupo C

Pode-se ver na figura 3.17 o gráfico de f_0 no ponto médio das vogais da palavra *colonoscopia*.

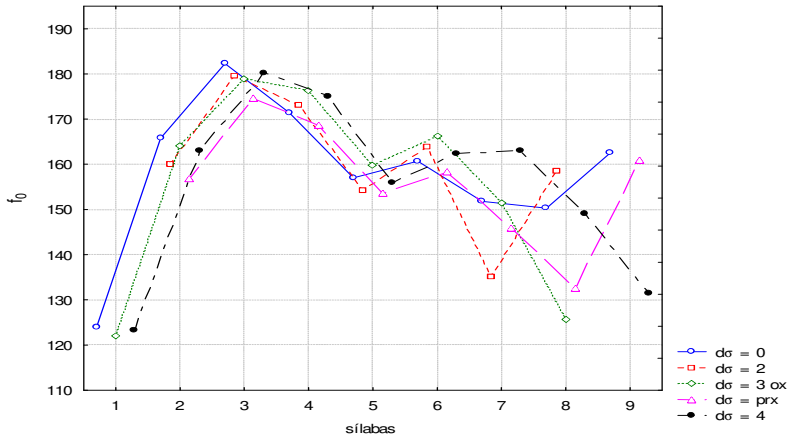


Figura 3.17: Médias de f_0 no ponto médio das vogais nas sentenças-veículo do grupo *C*. Os padrões de linhas representam os diferentes valores da variável d_σ.

Nos dados de f_0 no ponto médio fez-se uma ANOVAde duas entradas tendo como fatores PS (tomando neste caso as posições pré-tônicas[8]) e d_σ cujo resultado mostrou que o fator PS produz

[8]A tônica e a postônica foram realizadas como um ditongo, de modo que neste caso a tônica não funcionou como nos demais grupos.

efeito altamente signinifcativo ($F(3, 177) = 17,7\ p < 10^{-6}$), sendo não significativo o efeito do fator d_σ ($F(4, 177) = 0,9\ ns$), bem como a interação entre os dois ($F(12, 117) = 0,1\ ns$). As comparações *post hoc* no fator PS mostram que as posições 2 e 5 não diferem significativamente, assim como as posições 3 e 4. As comparações envolvendo os outros pares resultam sempre em diferenças significativas.

Grupo D

Pode-se ver na figura 3.18 o gráfico de f_0 no ponto médio das vogais da palavra *dirigibilidade*.

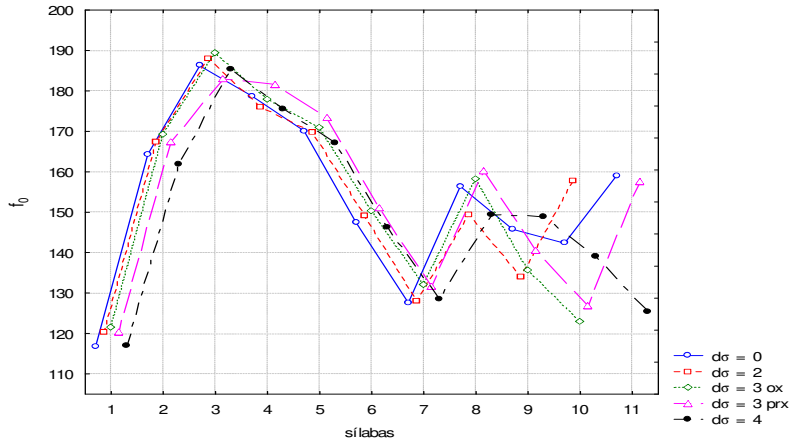


Figura 3.18: Médias de f_0 no ponto médio das vogais nas sentenças-veículo do grupo *D*. Os padrões de linhas representam os diferentes valores da variável d_σ.

Realizou-se uma ANOVA de duas entradas nos dados de f_0 no ponto médio tendo PS (posições 2–7) e d_σ como fatores. O primeiro mostrou-se significativo ($F(5, 258) = 66,7\ p < 10^{-6}$), o segundo e a interação entre ambos não atingiram o nível de significância — ($F(4, 258) = 0,6\ ns$) e ($F(20, 258) = 0,1\ ns$), respectivamente. No fator PS, os testes de comparação *post hoc* indicam que as únicas diferenças não significativas são as que envolvem as posições 2–4, 2–5, 3–4 e 4–5.

Demais posições

Quanto aos valores de f_0 no trecho que segue a palavra-chave nas sentenças-veículo não nos deteremos muito. Como os resultados referentes às sílabas pré-tônicas e tônicas mostraram, não há em nenhum dos grupos investigados influência substancial da variável d_σ. O efeito desta variável no parâmetro duração, ao contrário, mostrou-se significativo, no sentido de estar relacionado a uma maior ou menor duração da unidade do tamanho da sílaba lexicalmente acentuada de acordo com sua "distância" em relação à culminância do acento frasal. A variação do parâmetro f_0, portanto, não parece ser explicada pelo mesmo princípio que organiza a variação da duração, a saber, a relação entre acentuação e silabicidade. O resultado de adicionar palavras à direita da palavra-chave não parece alterar, como acontece no caso da duração, a maneira como o parâmetro f_0 se organiza ao longo do trecho do enunciado que compreende a palavra-chave, pelo menos nas sentenças estudadas.

O que se observa a partir de um exame das figuras de todos os grupos é que os valores dos pontos médios são muito semelhantes em todos os grupos. O que se observa é uma correlação entre o tipo acentual da palavra que segue a palavra-chave e o comportamento da f_0 no ponto médio.

3.4.2 Distribuição de pontos extremos H

A tabela 3.7 resume, para cada grupo de palavras no *corpus*, a porcentagem de ocorrência do número de extremos do tipo H, isto é, de picos na curva de f_0. A figura 3.19 mostra a distribuição relativa das ocorrências de extremos H e R ao longo das unidades V-V nos 4 grupos de palavras do *corpus*. As informações da tabela e da figura são complementares. A tabela informa as configurações tonais e sua frequência relativa e a figura mostra a distribuição espacial dos extremos H.

O exame da tabela mostra uma forte tendência à presença de dois Hs nas palavras-chave, a despeito do número de sílabas pré-tônicas. É curioso notar que a porcentagem de ocorrência de configurações monotonais, isto é, com apenas um H, aumenta à medida que a cadeia de pré-tônicas cresce, muito embora ela seja zero para a palavra com a maior cadeia de todas (*dirigibilidade*). De qualquer forma, a menor porcentagem de ocorrência de uma configuração bitonal é 76%, que é um número bastante alto.

2	3	4	5
1 H: 7%	1 H : 11%	1 H: 24%	2 Hs: 100%
2 Hs: 93%	2 H: 89%	2 H: 76%	

Tabela 3.7: Porcentagem de ocorrência do número de pontos extremos H por grupo de palavras com 2, 3, 4 ou 5 sílabas pré-tônicas.

O exame da figura mostra alguns fatos esperados. Um deles é que, em todos os grupos, as duas últimas unidades V-V tendem a concentrar ocorrências de pontos H. Como é comum acontecer com palavras paroxítonas a alinhamento dos extremos H pode flutuar em torno das duas últimas unidades, aquelas que contém a vogal tônica ou a pós-tônica. Note-se, entretanto, como os extremos R, isto é, os máximos da derivada da f_0, tendem a ter menor dispersão no seu alinhamento, concentrando-se na penúltima V-V, aquela que contém a tônica.

Uma análise possível para a distribuição dos H iniciais seria interpretá-los como manifestação de um acento secundário, pois nas palavras do grupo A (2 pré-tônicas) há bastante ocorrências de H na primeira sílaba da palavra-chave, o que estaria de acordo com a maioria das explicações para os acentos secundários. O mesmo valeria para o grupo B (3 pré-tônicas). Ha grande concentração de Hs na unidade V-V que corresponderia à segunda sílaba da palavra. Esse dado também se encaixaria nas previsões feitas pela hipótese de AS binários (acento secundário na sílaba *co* na palavra *locomotiva*).

Para a palavra *colonoscopia*, com 4 sílabas pré-tônicas, a hipótese de acentos secundários

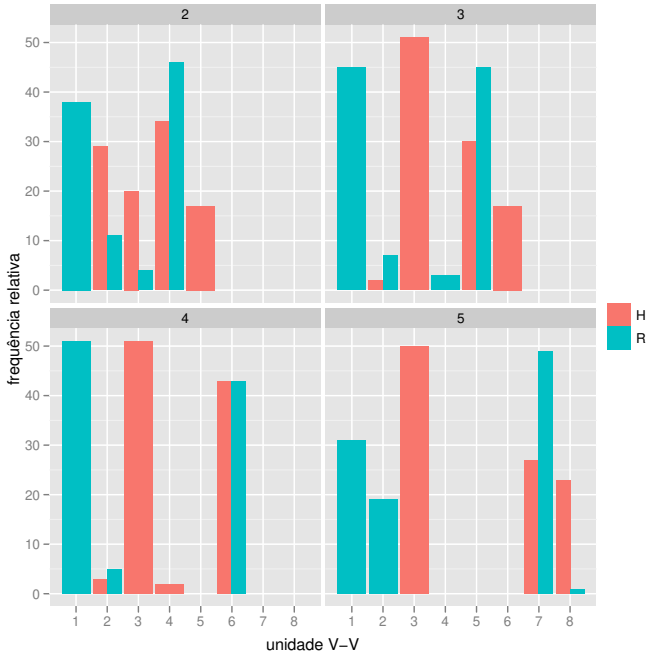


Figura 3.19: Frequência relativa da ocorrência de pontos extremos H e R por posição no grupo acentual nos grupos de palavras com 2, 3, 4 ou 5 sílabas pré-tônicas.

atribuídos de forma binária prevê dois acentos[9]: *còlonòscopia*. Entretanto, a figura mostra que a maioria dos Hs iniciais está concentrada na terceira posição V-V isto é, na segunda sílaba da palavra, o que resultaria na forma não prevista *colònoscopia*.

Para a palavra *dirigibilidade* seria possível sustentar que a grande concentração de ocorrências de H na posição 3 é resultado da atribuição binária, pois seria o primeiro dos dois secundários possíveis (*dirìgibìlidade*).

No caso dos grupos C e D é interessante notar que não ocorrem configurações com três Hs, que é uma configuração prevista pela hipótese dos AS atribuídos de forma binária.

Uma explicação alternativa que unificaria os padrões observados em todos os grupos na figura 3.19 seria dizer que em todos os grupos o que há é um acento tonal inicial, associado à ocorrência de extremos do tipo R na posição V-V 1 ou eventualmente 2. Do ponto de vista do processamento auditivo isso faria sentido, já que o ouvido humano é muito sensível a mudanças bruscas nos parâmetros acústicos. A presença de um R, que sinaliza um ponto de máxima variação na velocidade de mudança de f_0, na unidade V-V inicial indica que ali está ocorrendo um movimento ascendente na f_0. Nessa perspectiva, o alinhamento dos extremos H seriam a consequência superficial das mudanças na velocidade da f_0, cujos máximos são os pontos extremos do tipo R.

[9]Para fins de ilustração os AS serão notados pelo acento grave.

3.4.3 Discussão parcial

Os resultados apresentados parecem apontar para uma proeminência inicial marcada pelo parâmetro f_0 na forma de um pico localizado na maior parte dos casos na primeira sílaba da palavra-chave, podendo esta subida inicial, no caso do grupo B, estender-se também à segunda sílaba. Não parece plausível argumentar, ainda baseado naqueles resultados, em favor de uma alternância binária na produção do parâmetro f_0. Ainda que o resultado do grupo A possa sugerir uma intrepretação como esta, já que temos uma sequência alternante subida–descida–subida evidenciada pela taxa de variação de f_0 nas três primeiras sílabas das palavras daquele grupo, os resultados dos outros grupos desencorajam que se interprete o resultado do grupo A como sendo resultado da aplicação de um princípio de alternância binária. Ao contrário, fortalecem a interpretação de uma proeminência inicial, desde que a palavra tenha pelo menos duas sílabas pré-tônicas.

Uma outra possível interpretação, mais geral, para os resultados indicaria a formação de um padrão do tipo *hat pattern* recobrindo as pré-tônicas, seguido de outro recobrindo o conjunto tônica–pós-tônica. Mais uma vez, a interpretação desfavorece a postulação de algum tipo de alternância e aponta, como no caso da duração, para algum mecanismo que age no âmbito sintagmático mais largo do que, por exemplo, um possível pé métrico binário.

As sílabas pós-tônicas, na representação por pontos, geralmente têm valores médios maiores do que os verificados para as tônicas dado que poderia conduzir a uma análise em que a pós-tônica é considerada portadora de proeminência. A comparação dos dados da representação por pontos com os dados da análise por pontos extremos máximos e mínimos permite formular uma análise em que a sílaba tônica e a pós-tônica, pelo menos no caso das palavras paroxítonas, podem ser vistas, do ponto de vista da variação de f_0, como uma unidade. A observação da figura 3.19 mostra que as ocorrências dos extremos do tipo H estão distribuídas tanto na posição V-V correspondente à sílaba tônica quanto na posição correspondente à pós-tônica, enquanto as ocorrências dos extremos do tipo R estão distribuídas majoritariamente na posição correspondente à tônica. Essa diferença no comportamento dos dois tipos de extremo sugere que, apesar da variabilidade superficial do alinhamento de H, a associação mais sistemática dos extremos do tipo R com a sílaba tônica evidenciaria que é ela a portadora da proeminência devida ao movimento de f_0 e que a variabilidade do alinhamento dos Hs é consequência da possibilidade de um alinhamento mais tardio possibilitado pela existência de material fônico após a tônica.

3.5 Configuração formântica

3.5.1 Vogal /a/

Palavra *patarata*

A figura 3.20(a) mostra os valores médios de F_1 para a vogal /a/ nos cinco níveis da variável PS na amostra formada pelas repetições das frases-veículo contendo a palavra fonológica *A patarata*.

Fatores PS ($F(4, 210) = 219,\ p < 10^{-16}$) e d_σ ($F(4, 210) = 3.1,\ p < 0.02$) alcançaram

significância, mas não sua interação. Comparações pareadas mostram que todas as médias diferem entre si ($p < 10^{-3}$ ou menor), exceto pelas sílabas *pa* e *ra*. Nenhuma diferença entre os níveis da variável d_σ emergiu nas comparações pareadas.

A tabela 3.8 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e F_1. A célula da tabela com o maior valor de R^2 está assinalada em cinza. A figura 3.20(b) é um gráfico de dispersão que mostra os valores de F_1 como função da duração normalizada das sílabas. Na tabela e no gráfico os dados de F_1 são divididos em dois grupos, um contendo todas as posições sintagmáticas e outro em que a primeira posição é excluída. Essa decisão foi motivada pelo fato da inspeção dos gráficos de dispersão de F_1 em função das diversas modalidades de duração sugerir que o subgrupo composto pelos valores de F_1 da primeira posição (representados por círculos cheios na figura 3.20(b)) tem uma distribuição diferente daquela do grupo formado pelas demais posições. Comparando os coeficientes de determinação na tabela 3.8 pode-se observar que os valores são sistematicamente maiores no grupo em que a primeira posição é excluída.

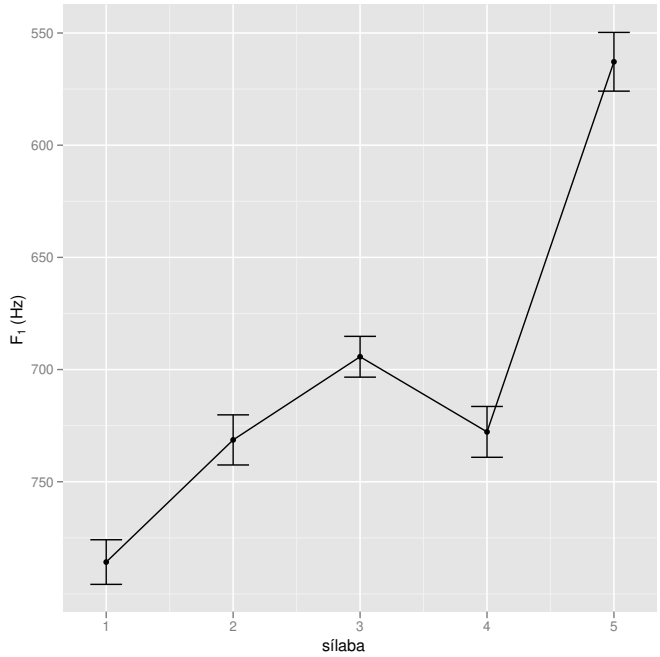
unidade	duração	posições	R^2	p
vogal	bruta	todas	0,04484	$< 0,01$
		1ª excluída	0,3027	$< 0,001$
	normalizada	todas	0,04483	$< 0,01$
		1ª excluída	0,3027	$< 0,001$
sílaba	bruta	todas	0,05219	$< 0,001$
		1ª excluída	0,2052	$< 0,001$
	normalizada	todas	0,02216	$< 0,05$
		1ª excluída	0,3854	$< 0,001$
unidade V-V	bruta	todas	0,1157	$< 0,001$
		1ª excluída	0,1701	$< 0,001$
	normalizada	todas	0,06362	$< 0,001$
		1ª excluída	0,1981	$< 0,001$

Tabela 3.8: Coeficientes de determinação (R^2) e nível de significância das correlações entre medidas de duração e F_1 em grupos acentuais contendo a palavra fonológica *A patarata*.

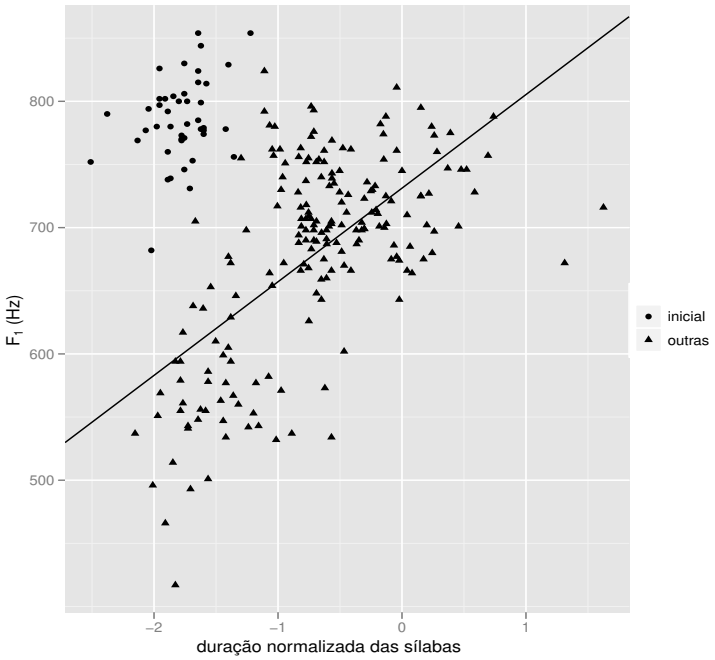
Palavra *jaratacaca*

A figura 3.21(a) mostra os valores médios de F_1 para a vogal /a/ nos seis níveis da variável PS na amostra formada pelas repetições das frases-veículo contendo a palavra fonológica *A jaratacaca*.

Fatores PS ($F(5, 232) = 148.3,\ p < 10^{-16}$) e d_σ ($(F(4, 193) = 4.4,\ p < 0.002)$) alcançaram significância, mas não sua interação. Comparações pareadas mostram que todas as médias para o fator PS diferem entre si ($p < 10^{-5}$ ou menor) exceto pela posição do artigo definido quando comparado à sílaba *ta* da palavra-alvo. Nenhuma diferença entre os níveis da variável d_σ emergiu nas comparações pareadas.



(a) Valores médios de F_1 para a vogal /a/ em cada vogal do grupo acentual. Barras verticais indicam intervalos de 95% de confiança em torno da média.



(b) Valores de F_1 como função da duração bruta das unidades V-V. A linha contínua é a reta de regressão linear com o melhor ajuste aos dados do subgrupo em que a primeira posição é excluída.

Figura 3.20

A tabela 3.9 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e F_1. A célula da tabela com o maior valor de R^2 está assinalada em cinza. A figura 3.21(b) é um gráfico de dispersão que mostra os valores de F_1 como função da duração bruta das unidades V-V.

unidade	duração	posições	R^2	p
vogal	bruta	todas	0,2724	$< 0,001$
		1ª excluída	0,3594	$< 0,001$
	normalizada	todas	0,2725	$< 0,01$
		1ª excluída	0,3595	$< 0,001$
sílaba	bruta	todas	0,02427	$< 0,05$
		1ª excluída	0,2208	$< 0,001$
	normalizada	todas	0,1501	$< 0,05$
		1ª excluída	0,2438	$< 0,001$
unidade V-V	bruta	todas	0,3268	$< 0,001$
		1ª excluída	0,3906	$< 0,001$
	normalizada	todas	0,2508	$< 0,001$
		1ª excluída	0,2973	$< 0,001$

Tabela 3.9: Coeficientes de determinação (R^2) e nível de significância das correlações entre medidas de duração e F_1 em grupos acentuais contendo a palavra fonológica *A jaratacaca*

3.5.2 Vogal /i/

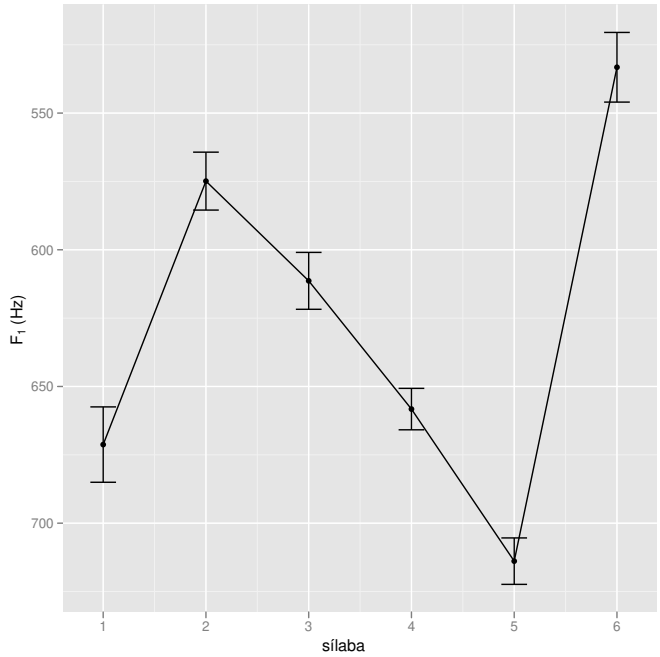
Uma ANOVA de duas entradas tendo as variáveis PS (restrita às posições 1–5 e 7) e d_σ como fatores aplicada aos dados de F_1 indicou ser a influência do primeiro fator altamente significativa ($F(5, 239) = 142,5\ p < 10^{-6}$) e não significativas as influências de d_σ e da interação entre os fatores — ($F(4, 239) = 1,3\ ns$) e ($F(20, 239) = 0,7\ ns$), respectivamente.

A aplicação dos testes de comparação *post hoc* ao fator PS indicou uma separação dos dados em três grupos: posições 2, 3 e 7, posição 1 e posições 4 e 5. Todas as comparações envolvendo elementos de um grupo com elementos de outro grupo resultam em diferenças significativas (com pelo menos $p < 10^{-5}$).

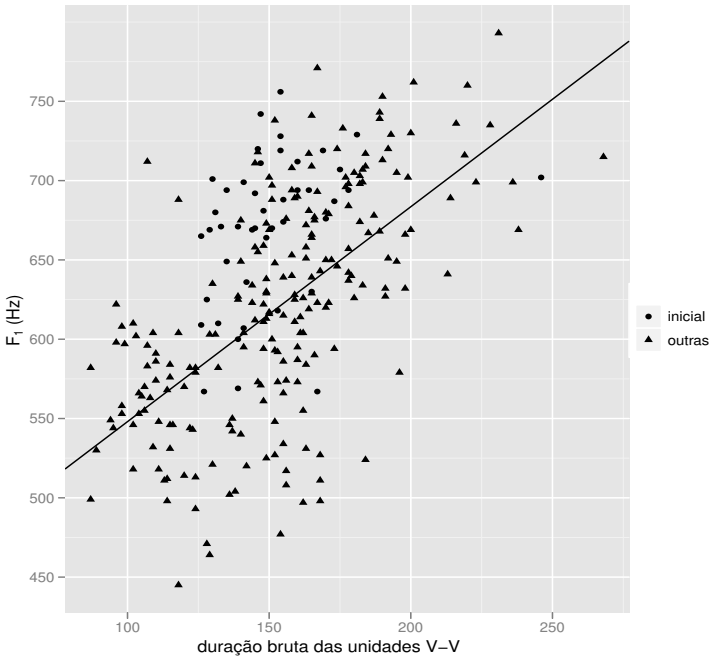
Estes resultados mostram que as posições 4 e 5, próximas à sílaba tônica, apresentam os maiores valores de F_1, indicando que as vogais nestas posições foram pronunciadas de forma relativamente mais aberta. As posições 2, 3 e 7 apresentaram os menores valores de F_1, indicando uma posição de mandíbula mais alta. A posição 1 apresenta valores intermediários entre as posições mais distantes (2 e 3) e as mais próximas da tônica (4 e 5). Há, portanto, uma gradiência na qualidade vocálica condicionada pela posição da vogal na palavra[10].

A figura 3.22 mostra os valores médios de F_1 da vogal /i/ em cada posição da palavra *dirigibilidade* (exceto a posição correspondente à vogal tônica /a/).

[10]Resultados semelhantes foram obtidos por Gama Rossi (1998), cf. seção 2.4.



(a) Valores médios de F_1 para a vogal /a/ em cada vogal do grupo acentual. Barras verticais indicam intervalos de 95% de confiança em torno da média.



(b) Valores de F_1 como função da duração bruta das unidades V-V. A linha contínua é a reta de regressão linear com o melhor ajuste aos dados do subgrupo em que a primeira posição é excluída.

Figura 3.21

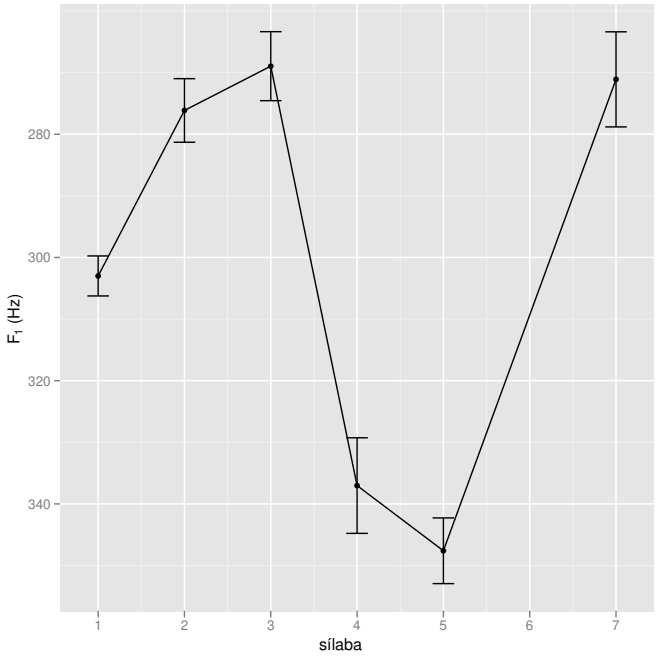


Figura 3.22: Valores médios de F_1 para a vogal /i/ em cada vogal do grupo acentual. Barras verticais indicam intervalos de 95% de confiança em torno da média.

3.5.3 Discussão parcial

Os resultados referentes aos dados de F_1 testemunham a existência de uma gradiência prosodicamente motivada na abertura vocálica. Os dados provenientes da análise dos grupos acentuais contendo a palavra fonológica *A patarata* sugerem que a primeira sílaba do grupo acentual (o artigo definido a) recebe proeminência relativa indicada por uma articulação vocálica relativamente mais aberta que se fecha nas duas vogais subsequentes, que correspondem às sílabas pré-tônicas do substantivo *patarata*. Uma abertura relativa volta a acontecer na vogal da sílaba tônica *ra* seguindo-se então um fechamento significativo na vogal pós-tônica, como era de se esperar. Este resultado poderia ser interpretado segundo a hipótese de que as sílabas pré-tônicas da palavra *patarata* são agrupadas em um pé métrico iâmbico. Esta explicação não dá conta da grande abertura vocálica observada para a primeira sílaba do grupo acentual (o artigo definido), que não seria deixado de fora na formação dos pés, segundo um dos algoritmos motivados por binaridade propostos na literatura (cf. Collischonn 1994) .

Os resultados da análise dos dados dos grupos acentuais contendo a palavra-alvo *jaratacaca* podem também ser interpretados como um caso de proeminência inicial sinalizada por uma abertura relativa da articulação da primeira vogal do grupo acentual. Diferentemente do caso discutido no parágrafo anterior, no entanto, a abertura relativa decresce abruptamente, de modo que uma diferença de pouco mais de 100 Hz em F_1 pode ser observada entre a primeira e a segunda vogais do grupo acentual. Um movimento contínuo e suave de abertura começa na segunda vogal e culmina na sílaba tônica do substantivo. Este padrão não poderia ser explicado por um algoritmo que constrói pés binários entre as pré-tônicas, uma vez que o

padrão observado é claramente de abertura gradiente.

Os dados da vogal /i/ na palavra *dirigibilidade* apontam na mesma direção dos resultados da vogal /a/. Observa-se igualmente uma abertura gradiente na articulação da vogal à medida que se aproxima o acento lexical frasalmente acentuado. A diferença em relação aos resultados da vogal /a/ é que a abertura relativa inicial ocorre na primeira vogal da palavra-chave. Como o artigo não foi medido no por ser uma vogal diferente das vogais das sílabas pré-tônicas da palavra *dirigibilidade*, não se pode afirmar se a vogal /a/ que corresponde ao artigo na palavra fonológica *A dirigibilidade* também apresentou uma realização mais aberta.

De maneira similar ao parâmetro duração, a configuração formântica das vogais também parece ser evidência para o caráter culminativo da construção das proeminências ao longo do grupo acentual. É importante também lembrar que a abertura diferenciada que recebe a primeira vogal da palavra é um resultado que se soma às evidências fornecidas pelos dados de duração e f_0 em favor de uma possível proeminência inicial, que se manifesta por meio de durações maiores na primeira unidade V-V das sentenças-veículo e grandes taxas de variação positiva de f_0 na primeira vogal das palavras-chave.

O modelo dinâmico do ritmo, muito embora faça previsões somente para o parâmetro acústico duração, pode dar pistas que ajudem a explicar em alguma medida dados de fortalecimento articulatório motivados por razões prosódicas. Existe uma relação entre duração de um fone (em especial as vogais) e abertura mandibular, uma vez que quanto maior for a excursão da mandíbula no sentido vertical (situação articulatória característica de uma vogal), maior será o intervalo de tempo gasto entre sair e voltar de uma posição mais fechada (situação articulatória característica das consoantes). Como o modelo prevê que o mecanismo de culminação induz o aumento progressivo das durações ao longo do grupo acentual a consequência esperada é, portanto, um aumento crescente da abertura mandibular no mesmo domínio, levando em conta, evidentemente, a natureza das vogais em jogo. Tanto na caso da palavra *patarata* quanto no da palavra *jaratacaca* os resultados das regressões lineares constituem evidência favorável a uma relativa dependência da abertura vocálica em relação à duração. De forma geral, o valor de R^2 diminui quando a posição 1 é incluída na amostra. Como se pode ver nas figuras 3.20(a) e 3.21(a), a vogal da posição inicial (o artigo definido que antecede a palavra-chave) tem valores de F_1 altos, indicando uma articulação mais aberta. Como as correlações entre duração e F_1 são todas positivas, indicando que quanto maior é a duração das vogais, sílabas ou unidades V-V maior será o valor de F_1, o fato dos R^2 piorarem como a inclusão dos valores da primeira posição indicam que o abertura vocálica observada na posição inicial não está fortemente relacionada com o alongamento duracional quer da vogal, sílaba ou unidade V-V inicial. Para as demais posições pode-se dizer que o grau de abertura vocálica está mais diretamente relacionado com o contorno duracional.

3.6 Resultados: ênfase espectral

3.6.1 Palavra *patarata*

A figura 3.20(a) mostra diferenças de energia entre bandas (em dB) para a vogal /a/ nas cinco sílabas do grupo acentual *A patarata*. Não serão discutidos os resultados das análises

estatísticas para a banda $B3-2$ porque para essa banda nenhum padrão significativo do ponto de vista prosódico pôde ser depreendido. Os valores para a diferença entre essas duas bandas serão omitidos das figuras também.

unidade	duração	posições	R^2	p
vogal	bruta	todas	0,2684	$< 0,001$
		1ª excluída	0,2896	$< 0,001$
	normalizada	todas	0,2684	$< 0,001$
		1ª excluída	0,2896	$< 0,001$
sílaba	bruta	todas	0,1298	$< 0,001$
		1ª excluída	0,2937	$< 0,001$
	normalizada	todas	0,2872	$< 0,001$
		1ª excluída	0,3583	$< 0,001$
unidade V-V	bruta	todas	0,1158	$< 0,001$
		1ª excluída	0,1159	$< 0,001$
	normalizada	todas	0,2141	$< 0,001$
		1ª excluída	0,2117	$< 0,001$

Tabela 3.10: Coeficientes de determinação (R^2) das correlações entre medidas de duração e ênfase espectral (*B2–B1*) em grupos acentuais contendo a palavra fonológica *a patarata*.

B2–B1: PS ($F(4, 171) = 70,\ p < 10^{-16}$) é o único fator que atinge significância. Comparações pareadas mostram as médias para o artigo definido e para a sílaba pós-tônica diferem significativamente ($p < 10^{-5}$ ou menos) das demais. As médias para as pré-tônicas e a tônica não diferem.

B4–B3: PS ($F(4, 210) = 57.7,\ p < 10^{-16}$) é o único fator que alcança significância. Comparações pareadas mostram que as médias para a sílaba pré-tônica *ta* e pós-tônica *ta* não diferem significativamente bem como as médias para o artigo definido *a* e a sílaba pré-tônica *pa*. Todas as demais comparações atingem significância ($p < 10^{-6}$ ou menor).

A tabela 3.10 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e os valores de ênfase espectral na banda *B2–B1*. A célula da tabela com o maior valor de R^2 está assinalada em cinza. A figura 3.20(b) é um gráfico de dispersão que mostra os valores de ênfase espectral nas bandas B2–B1 e B4–B3 como função da duração normalizada das sílabas. Na tabela e no gráfico os dados de F_1 são divididos em dois grupos, um contendo todas as posições sintagmáticas e outro em que a primeira posição é excluída.

A tabela 3.11 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e os valores de ênfase espectral na banda *B4–B3*. A células da tabela com o maior valor de R^2 está assinalada em cinza.

unidade	duração	posições	R^2	p

vogal	bruta	todas	0,1938	$< 0,001$
		1ª excluída	0,1426	$< 0,001$
	normalizada	todas	0,1939	$< 0,001$
		1ª excluída	0,1426	$< 0,001$
sílaba	bruta	todas	0,03355	$< 0,01$
		1ª excluída	0,03741	$< 0,01$
	normalizada	todas	0,065	$< 0,001$
		1ª excluída	0,01056	n.s.
unidade V-V	bruta	todas	0,03434	$< 0,01$
		1ª excluída	0,04289	$< 0,01$
	normalizada	todas	0,1184	$< 0,001$
		1ª excluída	0,1052	$< 0,001$

Tabela 3.11: Coeficientes de determinação (R^2) das correlações entre medidas de duração e ênfase espectral (*B4–B3*) em grupos acentuais contendo a palavra fonológica *a patarata*.

3.6.2 Palavra *jaratacaca*

A figura 3.24(a) mostra diferenças de energia entre bandas (em dB) para a vogal /a/ nas seis sílabas do grupo acentual *A jaratacaca*. Não serão discutidos os resultados das análises estatísticas para a banda $B3-2$ porque para essa banda específica nenhum padrão significativo do ponto de vista prosódico pôde ser depreendido. Os valores para a diferença entre essas duas bandas serão omitidos das figuras também.

B2–B1: PS ($F(5,232) = 93.8,\ p < 10^{-16}$) e d_σ ($F(4,193) = 2.6,\ p < 0.005$) alcançam significância, mas não sua interação. Comparações pareadas mostram que todas as médias diferem entre si ($p < 0.003$ ou menos) exceto pelas sílabas *ra* comparada à sílaba *ta*, bem como *ta* comparada à tônica *ca*. Nenhuma diferença significativa emergiu para o fator d_σ nas comparações pareadas.

B4–B3: PS ($F(5,232) = 48.5,\ p < 10^{-16}$ é o único fator que gera significância. A interação entre PS e d_σ é marginalmente significativa ($F(20,232) = 1.5,\ p = 0.07$). Comparações pareadas mostram significância para todos os pares de médias ($p < 0.005$ ou menos), exceto para a comparação entre o artigo definido *a* e a pós-tônica *ca*, entre *ja* e *ra* e, finalmente, entre a tônica *ca* e a pós-tônica *ca*.

A tabela 3.12 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e os valores de ênfase espectral na banda *B2–B1*. A célula da tabela com o maior valor de R^2 está assinalada em cinza. A figura 3.21(b) é um gráfico de dispersão que mostra os valores de ênfase espectral nas bandas B2–B1 e B4–B3 como função da duração normalizada das sílabas. Na tabela e no gráfico os dados de F_1 são divididos em dois grupos, um contendo todas as posições sintagmáticas e outro em que a primeira posição é excluída.

unidade	duração	posições	R^2	p

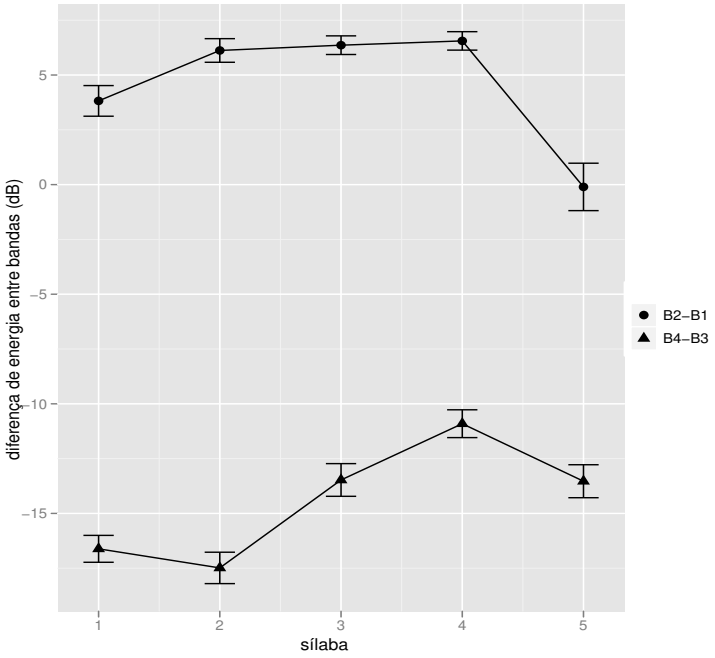
vogal	bruta	todas	0,2758	$< 0,001$
		1ª excluída	0,3044	$< 0,001$
	normalizada	todas	0,2758	$< 0,001$
		1ª excluída	0,3044	$< 0,001$
sílaba	bruta	todas	0,1928	$< 0,001$
		1ª excluída	0,08029	$< 0,001$
	normalizada	todas	0,2165	$< 0,001$
		1ª excluída	0,1783	$< 0,001$
unidade V-V	bruta	todas	0,194	$< 0,001$
		1ª excluída	0,205	$< 0,001$
	normalizada	todas	0,1617	$< 0,001$
		1ª excluída	0,1788	$< 0,001$

Tabela 3.12: Coeficientes de determinação (R^2) das correlações entre medidas de duração e ênfase espectral (*B2–B1*) em grupos acentuais contendo a palavra fonológica *a jaratacaca*.

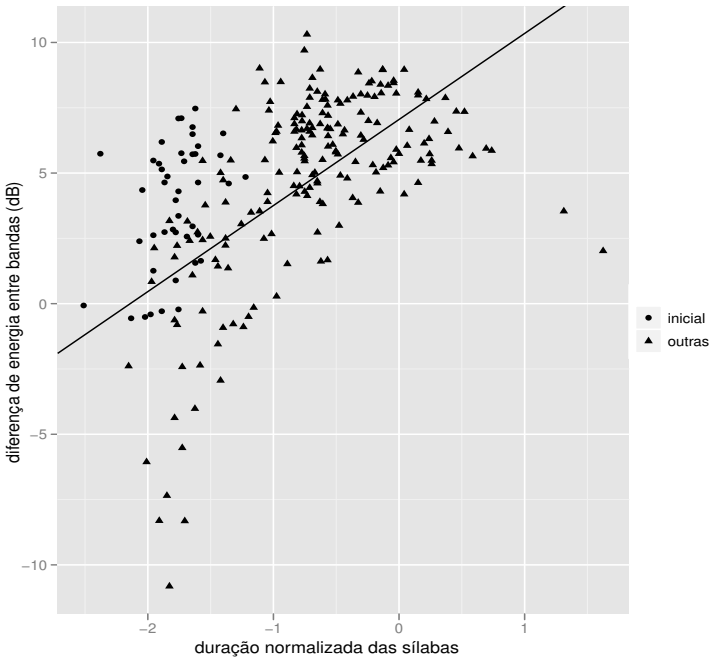
A tabela 3.13 mostra as correlações entre a duração de vogais, sílabas e unidades V-V e os valores de ênfase espectral na banda *B4–B3*. As células da tabela com os maiores valores de R^2 está assinalada em cinza.

unidade	duração	posições	R^2	p
vogal	bruta	todas	0,01026	n.s.
		1ª excluída	0,00454	n.s.
	normalizada	todas	0,01027	n.s.
		1ª excluída	0,0045	n.s.
sílaba	bruta	todas	0,0683	$< 0,001$
		1ª excluída	0,072	$< 0,001$
	normalizada	todas	0,003	n.s.
		1ª excluída	0,00015	n.s.
unidade V-V	bruta	todas	0,1485	$< 0,001$
		1ª excluída	0,1491	$< 0,001$
	normalizada	todas	0,0583	$< 0,001$
		1ª excluída	0,05472	$< 0,001$

Tabela 3.13: Coeficientes de determinação (R^2) das correlações entre medidas de duração e ênfase espectral (*B4–B3*) em grupos acentuais contendo a palavra fonológica *a jaratacaca*.

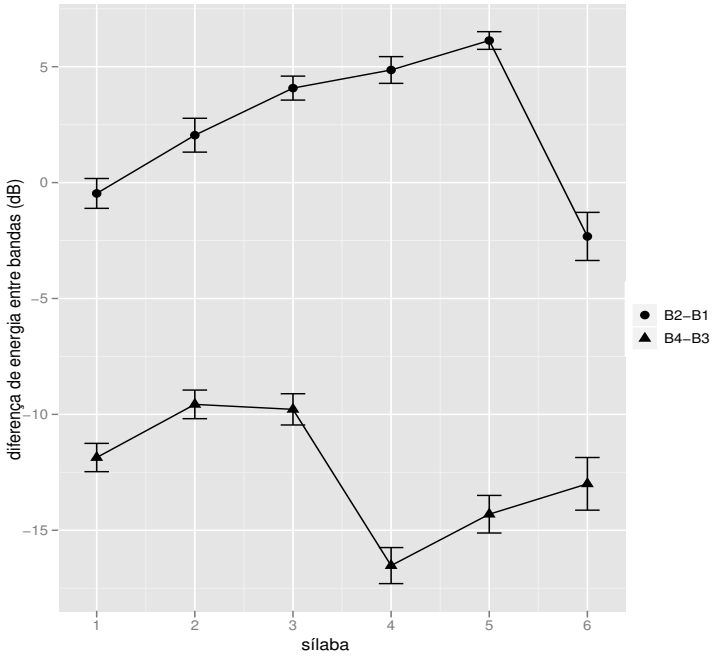


(a) Valores médios para a diferença de energia entre as bandas *B2–B1* e *B4–B3* para a vogal /a/ em cada sílaba do grupo acentual. Barras verticais indicam intervalos de 95% de confiança em torno da média.

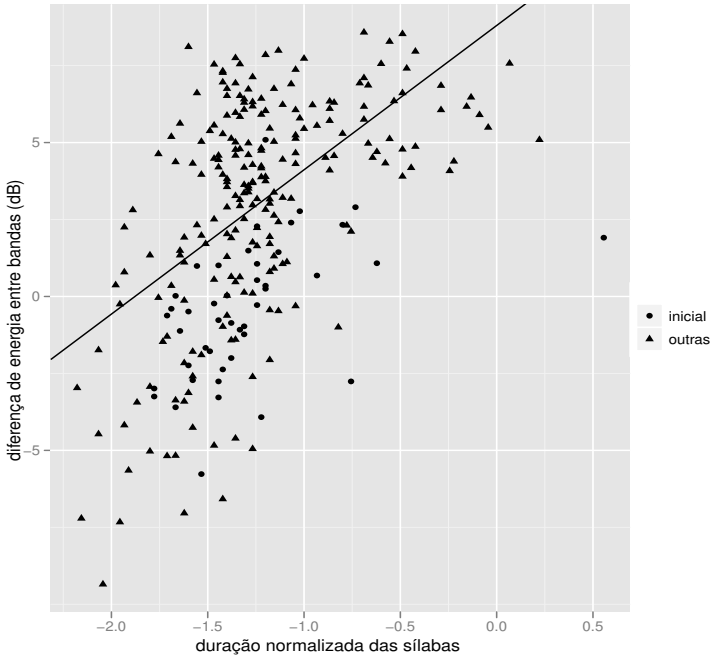


(b) Valores de ênfase espectral na banda *B2–B1* como função da duração normalizada das sílabas. A linha contínua é a reta de regressão linear com o melhor ajuste aos dados do subgrupo em que a primeira posição é excluída.

Figura 3.23



(a) Valores médios para a diferença de energia entre as bandas *B2–B1* e *B4–B3* para a vogal /a/ em cada sílaba do grupo acentual. Barras verticais indicam intervalos de 95% de confiança em torno da média.



(b) Valores de ênfase espectral na banda *B2–B1* como função da duração normalizada das sílabas. A linha contínua é a reta de regressão linear com o melhor ajuste aos dados do subgrupo em que a primeira posição é excluída.

Figura 3.24

3.6.3 Discussão parcial

Os resultados para o parâmetro ênfase espectral são diferentes para as duas palavras-alvo. Para grupos acentuais contendo a palavra *patarata*, B2–1 parece traçar uma espécie de linha divisória separando a sílaba pós-tônica do substantivo e o artigo definido das demais sílabas numa espécie de padrão de figura e fundo que dá destaque às sílabas pré-tônicas e tônica do substantivo (fazendo as vezes de figura) contra um pano de fundo menos proeminente. *B4–B3*, por outro lado, parece mostrar o desenrolar de um acento frasal culminativo que atinge seu ápice na sílaba tônica do substantivo. Isso significa que o envelope de energia das vogais passa a ter um caimento menos abrupto nas altas freqüências (mais energia presente na região em torno de F_3 e F_4 de /a/), sob a influência do mecanismo do acento frasal. *B3–B2* aparentemente não apresenta nenhum tipo de variação prosodicamente motivado.

A análise dos grupos acentuais contendo a palavra-alvo *jaratacaca* sugere que *B2–B1* serve ao mesmo tempo dois propósitos distintos: singularizar a vogal pós-tônica contra todas as demais e ao mesmo tempo refletir o desenrolar do acento frasal. Pode-se hipotetizar que o gradual reforço de energia em torno da região de F_1 poderia garantir maior inteligibilidade e proeminência para a sílaba tônica que coincide com o acento frasal. As diferenças de energia entre as demais bandas no caso dos grupos acentuais contendo essa palavra-alvo não parecem ter uma motivação clara.

Os resultados das análises de correlação entre duração e ênfase espectral mostram que a banda *B2–B1* é a que mais dependência apresenta em relação à duração. Os resultados mostram ainda que a exclusão da primeira posição não aumenta os valores de R^2 de forma tão sistemática quanto no caso dos resultados de qualidade vocálica, o que é possivelmente reflexo do fato de que o padrão dos dados de ênfase espectral não sugerem uma ênfase relativamente maior da primeira posição.

De forma geral, parece ser possível afirmar que há evidência preliminar que sugere que tanto a abertura vocálica quanto a ênfase espectral parecem seguir um padrão de culminância em parte semelhante ao observado para a duração, o que contribui com evidência favorável para a ideia de culminância que subjaz ao modelo de dinâmico do ritmo.

Capítulo 4

Estudo de produção II

Os proveitos analíticos representados pela possibilidade de entender os acentos secundários como proeminências prosódicas iniciais e pelo sucesso em sua reprodutibilidade por meio de simulações com o modelo dinâmico estão limitados ao universo coberto pelo *corpus* do experimento I que é composto pelos grupos acentuais em início absoluto de enunciado.

Malgrado o ganho conseguido pela escolha das variáveis independentes manipuladas no experimento de produção I, a posição da palavra-alvo na frase-veículo no *corpus* daquele experimento é fixa e esse fato limita o escopo de aplicação das inferências que se pode fazer a respeito do comportamento dos parâmetros acústicos investigados. O experimento descrito no presente capítulo foi planejado com o propósito tentar remediar essas limitações e conseguir um maior grau de generalidade descritiva para o fenômeno do acento secundário através da formulação de um novo *corpus*.

Na formulação do novo *corpus*, novos controles foram incorporados pela inclusão de variáveis independentes que inexistiam no corpus original. Essas variáveis são de natureza lexical, sintática e pragmático-discursiva. No nível lexical, além de palavras paroxítonas, também exemplares do padrão oxítono de acento foram acrescentadas ao conjunto das palavras-alvo. A frequência de uso também foi levada em consideração como variável lexical. No nível sintático, além da posição inicial, o corpus contém estruturas em que a palavra-alvo ocorre em posição medial e final na frase-veículo. No nível pragmático-discursivo, introduziu-se uma manipulação no *status* referencial do constituinte sintático em que a palavra-alvo está inserida. Por outro lado, uma vez que os resultados da análise dos parâmetros acústicos investigados no experimento de produção I não mostraram nenhuma influência significativa da natureza da vogal para além dos fenômenos da duração e f_0 intrínsecas, esse controle foi relaxado.

4.1 Variáveis lexicais

No *corpus* do experimento II a seleção das palavras-alvo adotou critérios diferentes em relação aos empregados no experimento anterior. A exigência de identidade entre todas as vogais na rima das pré-tônicas foi relaxada e admitiram-se palavras derivadas do ponto de vista morfológico. Os resultados da análise dos parâmetros acústicos não parecem mostrar, no experimento I, nenhuma influência prosódica relevante atribuível àqueles dois controles além

dos universais fenômenos de duração e f_0 intrínsecas. Incluir palavras derivadas no universo de busca permitiu escolher palavras mais frequentes e possivelmente mais familiares aos falantes. Outra novidade em relação ao experimento de produção I foi a inclusão de uma nova variável lexical. Para o experimento II foram escolhidas palavras com padrão acentual paroxítono e oxítono.

Aplicando os novos critérios de seleção, um conjunto de dezoito palavras-alvo foi delimitado para o novo *corpus*. Este total está distribuído entre palavras com padrão acentual paroxítono e oxítono e palavras com duas, três ou quatro sílabas pré-tônicas. No *corpus* novo procurou-se selecionar preferencialmente palavras com frequência de uso entre média e alta, de acordo com os critérios expostos mais adiante na seção 4.1.2. A tabela 4.1 lista as palavras selecionadas para o *corpus*.

pré-tônicas	acento	frequência	palavra
2	paroxítona	97,2	*delegado*
		266,8	*deputado*
		81,3	*jornalista*
	oxítona	239,6	*jogador*
		191	*militar*
		174,3	*senador*
3	paroxítona	229,9	*americano*
		109,9	*departamento*
		170,3	*secretaria*
	oxítona	126,8	*consumidor*
		271,3	*governador*
		59,8	*procurador*
4	paroxítona	56,6	*oportunidade*
		124,5	*possibilidade*
		215,1	*universidade*
	oxítona	3,5	*peregrinação*
		63,9	*privatização*
		6,5	*sofisticação*

Tabela 4.1: Lista de palavras-chave do *corpus* do experimento de produção II. A frequência de uso é listada em ocorrências por milhão.

O tamanho de uma palavra (medido em fones ou sílabas) e sua frequência de uso na fala ou em textos são duas grandezas que tendem a seguir a lei de Zipf (cf. figura 4.1 que ilustra dados reportados em Zipf 1965[1935], pp. 20–29, de um *corpus* de quase 11 milhões de ocorrência de alemão escrito), i.e., estabelecem entre si uma relação de proporcionalidade inversa, de modo que quanto maior é uma grandeza, menor é a outra. São mais raras, portanto, as palavras polissilábicas de alta frequência de uso.

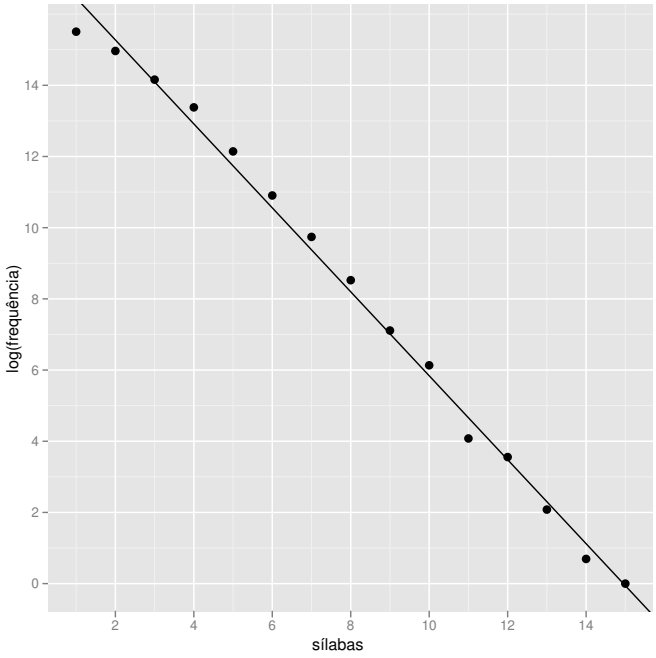


Figura 4.1: Logaritmo da frequência bruta de ocorrências de palavras com número crescente de sílabas. A linha contínua é a reta de regressão linear com o melhor ajuste aos dados.

4.1.1 Obtenção da frequência de uso

Para obter a informação de frequência de uso para guiar a escolha das palavras-alvo do corpus do experimento II recorreu-se a dois extratos do "Banco de Português", corpus de ocorrências de palavras organizado pelo projeto Direct, sediado na PUC–SP. No site do projeto[1] disponibiliza-se, em dois arquivos separados, dois extratos do "Banco de Português", um de fontes escritas e outro de fontes de fala, que trazem numa coluna uma lista dos tipos (*types*, em inglês) e em outra o respectivo número de ocorrências (*tokens*, em inglês). O extrato escrito contém em torno de 33 milhões de ocorrências de palavras (*tokens*), distribuídas em aproximadamente 255 mil ocorrências de tipo (*types*). O extrato de fala é composto de quase 3 milhões de *tokens* distribuídos entre pouco mais de 46 mil *types*. O extrato, segundo Sardinha (em comunicação por correio eletrônico) pode ser classificado, em termos de suas fontes, como um corpus generalista com predominância de escrita de imprensa.

4.1.2 Definindo faixas de frequência

Uma vez obtida a informação de frequência absoluta de uso para um conjunto de palavras candidatas torna-se necessário estabelecer uma métrica que permita posicionar uma palavra particular numa escala que discrimine regiões de alta ou baixa frequência de uso. Uma vez que diferentes autores usam diferentes critérios, resolveu-se seguir o uso comum no campo como uma

[1] Acessível através do endereço `http://www2.lael.pucsp.br/corpora/bp/`.

estratégia que facilita a comparabilidade dos resultados. A referência metodológica escolhida neste caso foi Bybee (1999 e 2001), por ser uma das grandes referências na área da fonologia do uso, que procura incorporar variáveis lexicais como a frequência de uso na teoria fonológica. Nos trabalhos citados, Bybee menciona o uso de dois corpora de referência para estimação de frequência de uso absoluta, um para experimentos com a língua inglesa (Francis e Kučera 1982) e outro para aqueles feitos com a língua espanhola (Juilland e Chang-Rodriguez 1964), ambos contendo em torno de um milhão de ocorrências (*tokens*). A autora adota duas categorias de frequência, baixa e alta, e estabelece como critério para classificação o limiar de 35 ocorrências por milhão no corpus de referência do inglês (Bybee 2001, p. 112) e 124 ocorrências por milhão para o corpus de referência do espanhol (Bybee 2001, pp. 148–9). Nenhum comentário é feito sobre a adoção de limiares diferentes para inglês e espanhol. Hammond (1999) usa um *corpus*[2] de 20 mil ocorrências do inglês para estimar a frequência de uso dos seus estímulos, mas não fornece nenhuma informação sobre os limiares adotados para a classificação em sua escala tripartite (baixa, média, alta). Não se relata ali também nenhum resultado de comparação estatística sistemática entre as três categorias de modo que não se pode dizer se a postulação de três categorias de frequência trazem ganhos em relação a uma escala binária (baixa e alta), por exemplo.

4.1.3 Frequência de uso e ritmo

Uma das razões para controlar a frequência de uso como variável independente é que existe evidência da influência desta variável na manifestação de fenômenos rítmicos. Um deles é a assim chamada regra do ritmo (*rhythm rule*, em inglês) cuja atuação seria responsável, por exemplo, pelo fenômeno do deslocamento acentual (*stress shift*, em inglês). Evidência favorável a essa proposição é relatada em Hammond (1999), que apresenta resultado de um experimento em que os sujeitos deveriam ler silenciosamente expressões nominais formadas por pares de adjetivo+substantivo (como *thirteen men*) e indicar, usando sua intuição linguística, "whether stress sounded better" (*ibidem*, p. 335) na primeira (situação que pode ser representado como *thírteen men*) ou na segunda sílaba (*thirtéen men*) do adjetivo ou se as duas opções lhes pareciam indiferentes. Nos estímulos elaborados por Hammond os substantivos são sempre monossilábicos de frequência média e os adjetivos, sempre bissilábicos, distribuem-se em duas categorias ortogonais: frequência (alta, média e baixa) e conformação morfológica (monomorfêmicos e derivados). Os resultados das preferências dos sujeitos mostram que entre as palavras morfologicamente simples a frequência de uso está diretamente correlacionada com a chance de ocorrência de deslocamento acentual motivado pela aplicação da regra do ritmo. Assim, dado um par com um adjetivo de baixa frequência como *oblique view* os sujeitos tendem a dizer que o acento "sounds better" na segunda sílaba do adjetivo (situação que Hammond representa por *oblíque view*) mais do que o contrário (*óblique view*). No caso das palavras morfologicamente complexas não se verifica um efeito significativo devido à variável frequência. De maneira geral, quando o adjetivo do par é derivado os sujeitos tendem a julgar como mais aceitável o acento na primeira sílaba malgrado a faixa de frequência[3]. A acomodação do resultado empírico que

[2] Acessível através da página do autor `http://www.lexicon.arizona.edu/~hammond/`

[3] Para uma crítica da noção de choque acentual, que motivaria a aplicação da regra do ritmo, ver as seção 3.2, 3.3 e 3.4 de Barbosa (2006, pp. 106–129), nas quais o autor aponta a aponta a improcedência dessa em

mostra ser a intuição dos falantes, tomada como representativa de sua competência linguística no sentido chomskyano, possivelmente afetada por uma variável lexical que tradicionalmente considera-se extra-linguística impõe um desafio para as explicações fonológicas para o fenômeno da aplicação da regra do ritmo que afirmam ser a intuição do falante alimentada apenas por informação acerca da distribuição dos acentos[4]. Uma vez que a maior parte das propostas em relação ao acento secundário propõe que sua atribuição seja guiada apenas por critérios estruturais, como a contagem do número de sílabas pré-tônicas, parece ser interessante investigar se fatores de outra ordem, como a frequência de uso têm alguma relevância em sua manifestação.

4.2 Variáveis sintáticas

No novo corpus cada palavra-alvo aparece em três frases-veículo diferentes, formuladas de modo a que ela ocupe três diferentes posições, conforme o exemplo 3 abaixo.

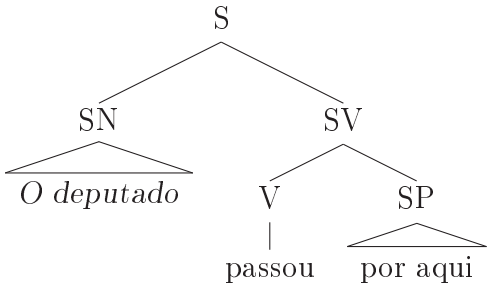
(3) a. *O deputado* passou por aqui hoje.

b. Pedro viu *o deputado* passar por aqui.

c. Paulo viu *o delegado*. Ele passou por aqui hoje.

No exemplo 3a, a palavra fonológica *o deputado* está em posição inicial, de modo semelhante às frases-veículo do corpus inicial (cf. em 4 a representação arbórea simplificada da oração), como no caso das frases do corpus inicial.

(4)

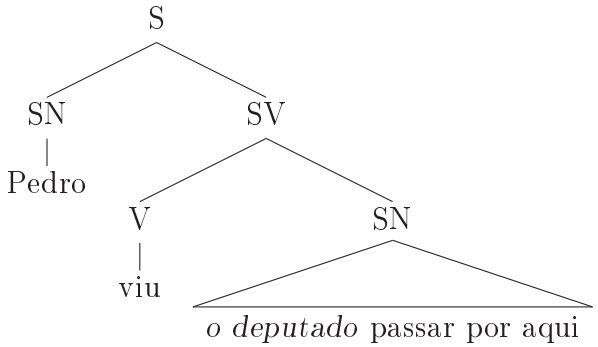


No exemplo 3b, a palavra fonológica *o deputado* inicia a oração subordinada substantiva objetiva direta reduzida de infinitivo – "o deputado passar por aqui" – estando em posição medial na oração completa, representado pela árvore sintática em 5.

função da falta evidência empírica que a apoie. Nas seções apontadas são apresentados dados fonéticos relativos ao inglês americano, PB e português europeu.

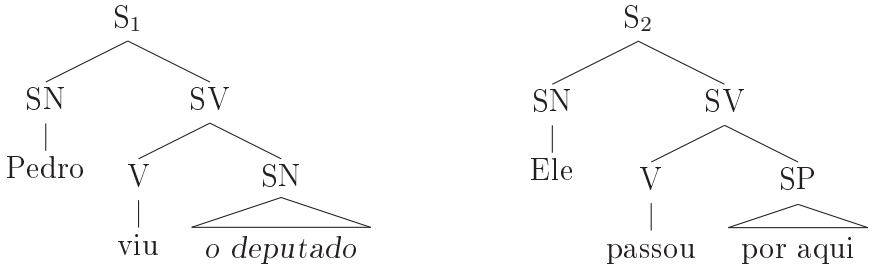
[4] O que não quer dizer que seja impossível, do ponto de vista técnico, que a teoria possa incorporar variáveis de natureza lexical. A questão é que essa incorporação correria o risco de se tornar um apêndice *ad hoc*.

(5)



Por fim, no exemplo 3c, *o deputado* está em posição de final de oração (cf. S_1 na árvore em 6). Evitou-se deixar a palavra-alvo em posição de final absoluto em função das dificuldades que os processos fônicos comuns a esta posição podem causar para a extração dos parâmetros acústicos que pretende-se estudar. Contudo, mesmo em posição final não-absoluta é possível observar fenômenos, como o resetting da f_0 e alongamento final, que podem exercer algum tipo de influência na marcação da acentuação secundária, o que justifica a inclusão desta variação sintática no novo corpus.

(6)



As variáveis sintáticas introduzidas no corpus novo podem também ser descritas em função das marcas relacionais de dependência estabelecidas por (Barbosa 2006, p. 226–7) para codificar diferentes graus de força de coesão sintática que liga constituintes contíguos (ou núcleos de dependência, nos termos de uma gramática de dependência). O autor mostra que há uma relação entre o grau de coesão existente entre dois constituintes e a probabilidade de ocorrência de uma fronteira prosódica entre os dois marcada pela realização de um acento frasal. No exemplo 7 mostra-se como as frases-veículo usadas como exemplo em 3 são etiquetadas segundo o sistema exposto em Barbosa (2006). Nos exemplos em 7, DFE quer dizer dependência forte à esquerda, DFD dependência forte à direita e IDF, por fim, independência forte.

(7) a. *O deputado* (DFE) passou por aqui hoje.
b. Pedro viu (DFD) *o deputado* (DFE) passar por aqui.
c. Paulo viu (DFD) *o delegado.* (IDF) Ele passou por aqui hoje.

4.3 Variável discursiva

Tendo em mente o objetivo de enriquecer a descrição do acento secundário por meio da cobertura da gama mais ampla possível de fatores que podem por hipótese ter algum tipo

de influência em sua manifestação, no desenho do novo corpus acrescentou-se aos controles expostos nas seções anteriores uma variável independente de natureza pragmático-discursiva.

Para a implementação desse controle foi criada uma situação em que a palavra-alvo aparece em duas orações adjacentes de forma que em cada ocorrência a palavra-alvo tenha *status* referencial distinto. No exemplo 8, a primeira ocorrência de *deputado* introduz um referente novo no discurso e a segunda ocorrência carrega o *status* de referente dado.

(8) Um *deputado* foi depor no plenário da Câmara Federal. O *deputado* foi evasivo em suas declarações.

Além da repetição da palavra-alvo, gerada pela justaposição das orações, a classificação de *status* dos dois referentes é reforçada pelo uso do artigo indefinido na função de determinante no sintagma nominal em *Um deputado* na primeira oração seguido do uso do artigo definido no sintagma nominal *O deputado* na segunda oração. Tradicionalmente, o artigo indefinido é considerado um introdutor de referentes novos no universo do discurso, ao passo que o definido é reconhecido como um recurso de retomada de referentes previamente mencionados.

Existem escalas que classificam de forma muito mais refinda os graus de novidade ou antiguidade (*giveness*) de referentes, como Prince (1981) e Gundel, Hedberg e Zacharski 1993). Estudos mostram, entretanto, que parte dos graus intermediários tendem a ser expressos mais comumente por meio de recursos sintáticos do que prosódicos (Brown 1983). Para uma classificação dicotômica, como a adotada aqui, por outro lado, existem dados mostrando que a prosódia pode ter um papel na diferenciação do *status* referencial (cf. Wolters 1999 e referências citadas pela autora para uma lista não-exaustiva). No nível prosódico, um dos possíveis correlatos do *status* referencial, segundo relatam Fowler e Housum (1987), é o encurtamento da duração das segundas ocorrências ou repetições de palavras em contextos conversacionais semi-espontâneos (o *corpus* utilizado pelos autores é a fala de um apresentador de televisão). No *corpus* do experimento a repetição é um dos recursos utilizados para mudar o *status* referencial.

Resultados como os de Fowler e Housum (1987), indicativos de uma influência da organização do discurso na estruturação temporal da fala, sugerem ser importante investigar a influência de índices de organização discursiva na manifestação de fenômenos como o acento secundário. Além disso, há muito poucos dados coletados em PB a respeito da relação entre prosódia e *status* referencial. Dada a importância deste tema para áreas de pesquisa como a pragmática, a semântica e mesmo para a síntese de fala, isso parece constituir ainda mais razão para a inclusão deste novo controle no corpus novo.

A lista completa das frases-veículo encontra-se no apêndice C

4.4 Metodologia

4.4.1 Procedimentos de gravação

Foram obtidas cinco repetições lidas de cada frase-veículo, gravadas em uma sala silenciosa no IEL/Unicamp por um locutor masculino adulto (em torno de trinta anos na época das gravações), falante nativo da cidade de Piracicaba-SP. O sinal acústico foi gravado diretamente

em formato digital pela placa de som MBox2 da Digidesign a uma taxa amostral de 44,1 kHz. Utilizou-se um microfone unidirecional Behringer modelo B-2 Pro para a captação. O *corpus* definido para o projeto é composto por dezoito palavras-chave, cada uma delas sendo inserida em 5 frases-veículo, sendo cada combinação possível repetida cinco vezes pelo sujeito informante, totalizando 360 frases gravadas.

4.4.2 Variáveis dependentes

Os parâmetros acústicos analisados no *corpus* do experimento de produção II como variáveis dependentes são a duração e a frequência fundamental. As análises avaliaram os efeitos do tipo acentual, número de sílabas pré-tônicas e das variáveis sintática e pragmático-discursiva sobre a duração das sílabas e unidades V-V, no caso da duração e sobre os padrões de distribuição de extremos altos (H) e ascendentes (R) sobre a variação global, no caso da frequência fundamental.

4.5 Resultados: duração

Para as finalidades da análise estatística as palavras-chave serão divididas em seis grupos segundo a combinação de dois critérios: (a) tipo acentual (**P**aroxítonas e **O**xítonas) e (b) número de sílabas pré-tônicas (**2**, **3** ou **4**). Para fazer referência aos seis grupos serão usados os códigos P2, P3, P4, O2, O3 e O4. Cada grupo é formado por três palavras (cf. tabela 4.1, p. 84).

Nas próximas seções serão analisados os efeitos da variável sintática (seção 4.5.1 e da variável discursiva (seção 4.5.2) sobre a duração de sílabas e unidades V-V. A variável PS (número de posições no sintagma) tem um número diferente para cada grupo de palavras. A tabela 4.2 mostra o número de posições e as palavras que compõem cada grupo.

grupo	unidades V-V	palavras
P2	5	*delegado* *deputado* *jornalista*
O2	4	*jogador* *militar* *senador*
P3	6	*americano* *departamento* *secretaria*
O3	5	*consumidor* *governador* *procurador*
P4	7	*oportunidade* *possibilidade* *universidade*

O4	6	*peregrinação* *privatização* *sofisticação*

Tabela 4.2: Número de unidades V-V analisadas em cada grupo de palavras-chave no experimento de produção II.

O grupo das palavras paroxítonas tem sistematicamente uma posição a mais em relação às oxítonas em função da pós-tônica. O agrupamento em unidades V-V seguiu os mesmos critérios expostos na seção 3.1.1. No agrupamento em sílabas incluiu-se também o artigo que precede a palavra-chave nas frases-veículo.

Na análise estatística dos contornos de duração as comparações pareadas em busca de padrões de proeminência relativa limitaram-se às posições iniciais, tanto no caso das unidades V-V quanto no caso das sílabas.

4.5.1 Efeito da variável sintática

Nesta seção serão analisados os efeitos da variável sintática sobre os contornos de duração de unidades V-V e sílabas em cada um dos seis grupos de palavras descritos acima. Cada nível da variável sintática ('inicial', 'medial' e 'final') foi analisado separadamente. Para cada nível aplicaram-se duas ANOVAs de um fator tendo como variável independente as posições sintagmáticas e como variável dependente a duração normalizada das unidades V-V ou das sílabas.

No nível 'final' da variável sintática há uma posição a menos em relação aos demais níveis nos dados referentes às unidades V-V. A título de exemplo, na frase "Pedro viu o deputado passar...", em que a palavra-chave *deputado* está em posição medial, o trecho a ser analisado é composto de cinco unidades V-V[5]: |iu o d|ep|ut|ad|o p|. Na frase "Paulo viu o deputado. Ele ...", em que a palavra-chave *deputado* está em posição final delimitam-se apenas quatro unidades: |iu o d|ep|ut|ad|. Neste último exemplo, delimitar uma quinta unidade V-V implicaria, segundo a definição adotada neste trabalho, incluir no cômputo da duração dessa unidade, além da vogal pós-tônica da palavra-chave, também a pausa silenciosa entre as duas frases. Optou-se então desprezar a última vogal uma vez que a unidade V-V que se formaria nestas condições seria a única a incluir uma pausa. O mesmo critério foi utilizado para as palavras oxítonas, embora neste caso a vogal ignorada pela delimitação seja a vogal da sílaba tônica. A decisão de excluir estas unidades não deve causar problema porque o objetivo principal da análise dos contornos de duração é verificar a possível existência de padrões de proeminências entre as pré-tônicas.

[5]Nos exemplos deste parágrafo respeitou-se a ortografia e a separação entre palavras para facilitar a visualização.

Grupo P2

A figura 4.2 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

- Unidades V-V

  Inicial $F(4, 70) = 29,8\ p < 0,001$; posições 1 e 2 ($p < 0,05$), 2 e 3 (n.s.), 1 e 3 ($p < 0,05$)

  Medial $F(4, 70) = 31\ p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 (n.s.), 1 e 3 ($p < 0,001$)

  Final $F(3, 52) = 25\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 1 e 3 (n.s.)

- Sílabas

  Inicial $F(4, 70) = 66,6\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 1 e 3 ($p < 0,001$)

  Medial $F(4, 70) = 60,9\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 1 e 3 ($p < 0,001$)

  Final $F(4, 65) = 29,7\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 1 e 3 (n.s.)

Grupo O2

A figura 4.3 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

- Unidades V-V

  Inicial $F(3, 56) = 54,2\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 1 e 3 ($p, 0,001$)

  Medial $F(3, 55) = 34,7\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,001$), 1 e 3 ($p < 0,001$)

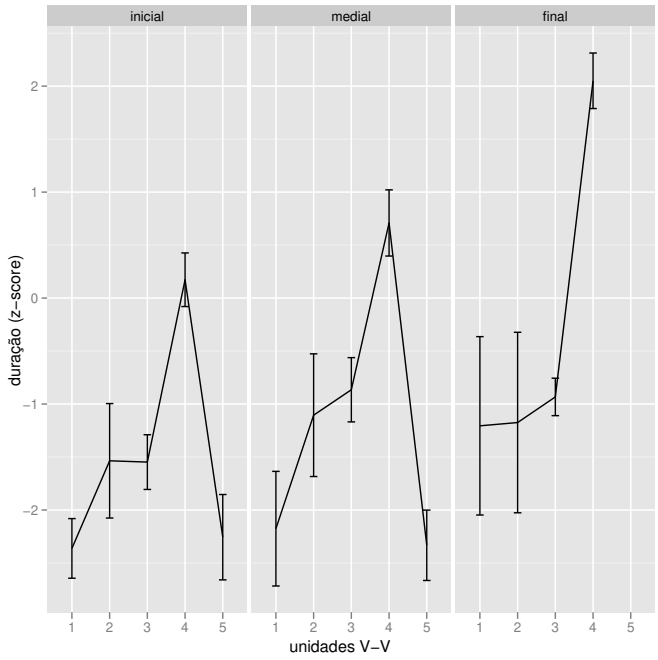
  Final $F(2, 42) = 30,2\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,001$), 1 e 3 (n.s.)

- Sílabas

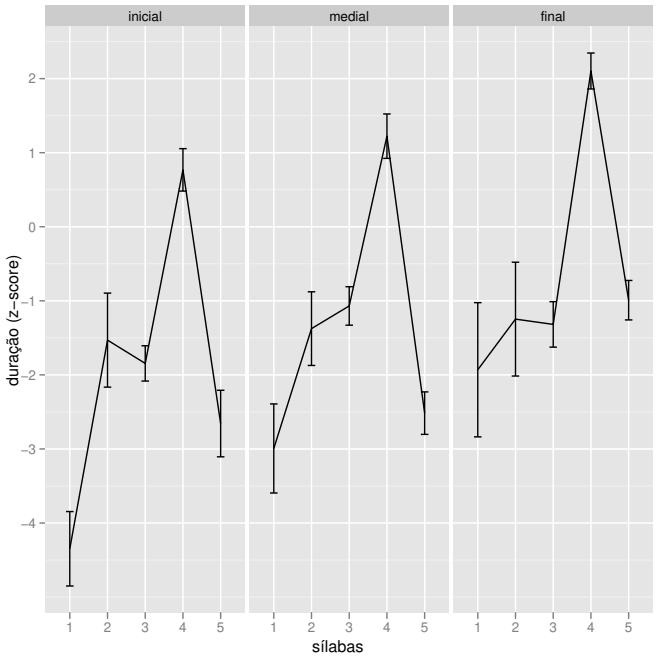
  Inicial $F(3, 56) = 112,8\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,05$), 1 e 3 ($p < 0,001$)

  Medial $F(3, 56) = 46,2\ p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 (n.s.), 1 e 3 ($p < 0,001$)

  Final $F(3, 56) = 112,5\ p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 (n.s.), 1 e 3 ($p < 0,001$)

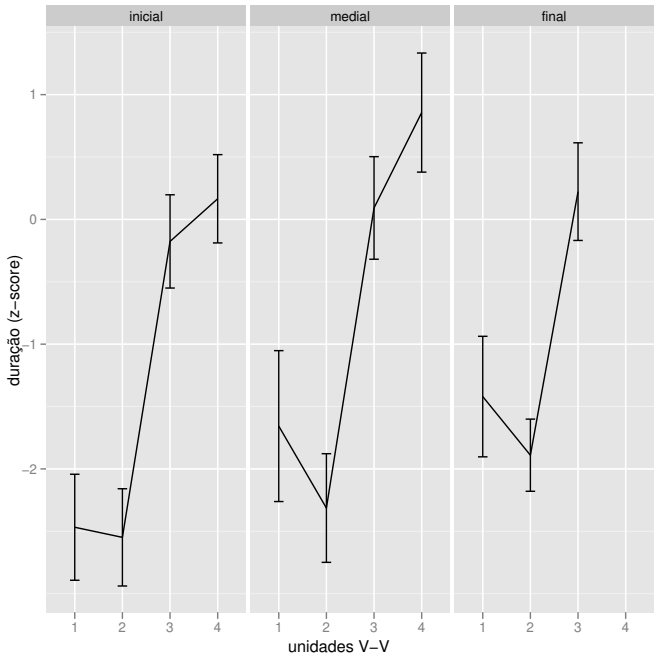


(a) Unidades V-V

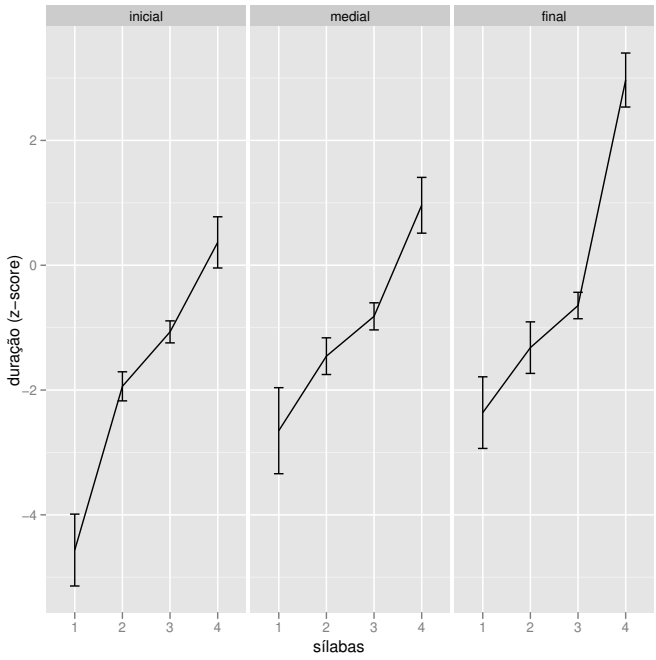


(b) Sílabas

Figura 4.2: Duração normalizada de unidades V-V e sílabas das palavras do grupo P2



(a) Unidades V-V



(b) Sílabas

Figura 4.3: Duração normalizada de unidades V-V e sílabas das palavras do grupo O2

Grupo P3

A figura 4.4 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

- Unidades V-V

  Inicial $F(5, 79) = 14,5\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

  Medial $F(5, 74) = 9,1\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0.01$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

  Final $F(4, 65) = 11,8\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,1$), 1 e 3 (n.s.), 1 e 4 (n.s.), 2 e 4 ($p < 0,05$)

- Sílabas

  Inicial $F(5, 79) = 15,9\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 ($p < 0,05$), 1 e 4 ($p < 0,001$), 2 e 4 (n.s.)

  Medial $F(5, 72) = 11,2\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 ($p < 0,05$), 2 e 4 (n.s.)

  Final $F(5, 74) = 14,9\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 2 e 4 (n.s.)

Grupo O3

A figura 4.5 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

- Unidades V-V

  Inicial $F(4, 66) = 45\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

  Medial $F(4, 69) = 52,7\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 ($p < 0,01$), 1 e 4 (n.s.), 2 e 4 ($p < 0,1$)

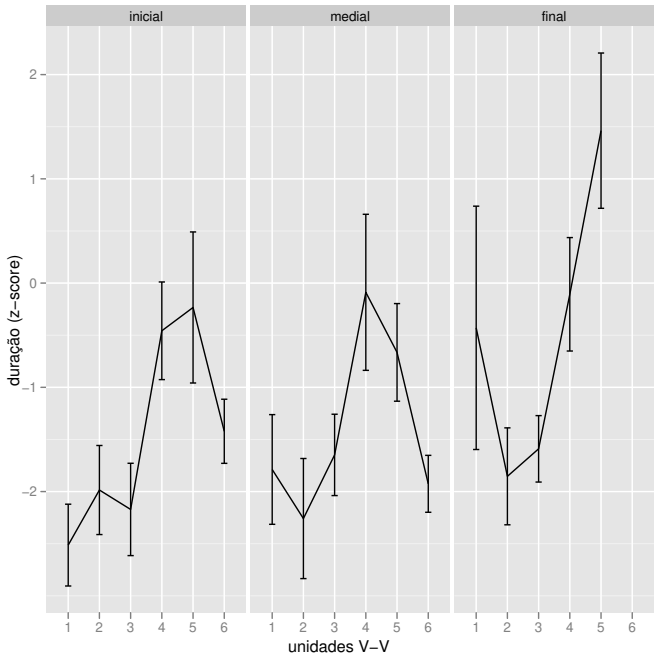
  Final $F(3, 56) = 18\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,01$), 3 e 4 ($p < 0,001$)

- Sílabas

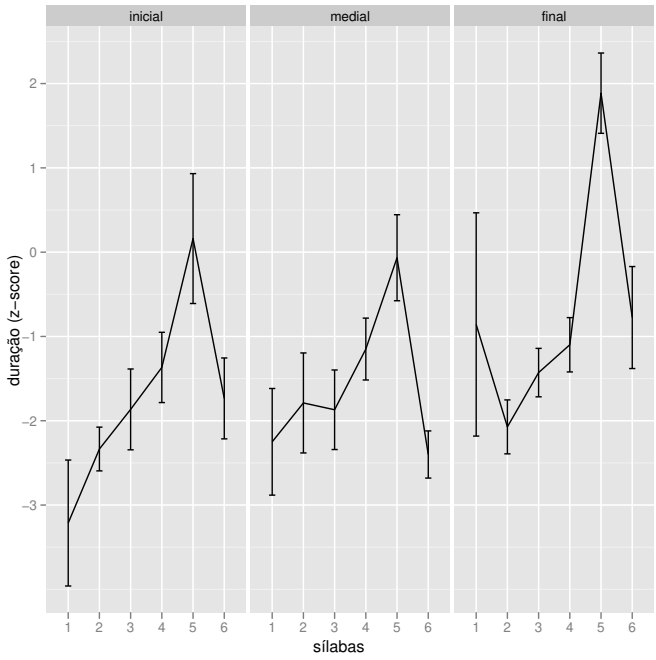
  Inicial $F(5, 79) = 15,9\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 ($p < 0,05$), 1 e 4 ($p < 0,001$), 2 e 4 (n.s.)

  Medial $F(5, 72) = 11,1\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 ($p < 0,05$), 2 e 4 (n.s.)

  Final $F(5, 74) = 14,8\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 2 e 4 (n.s.)

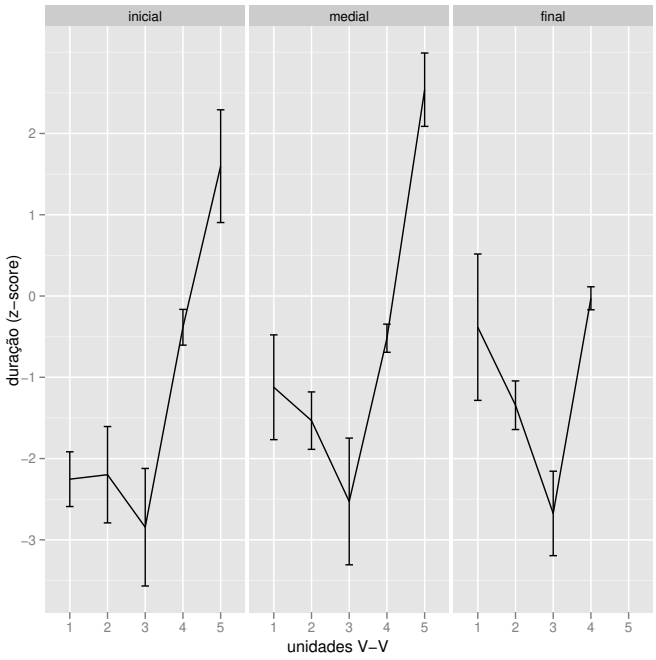


(a) Duração das unidades V-V das palavras do grupo P3

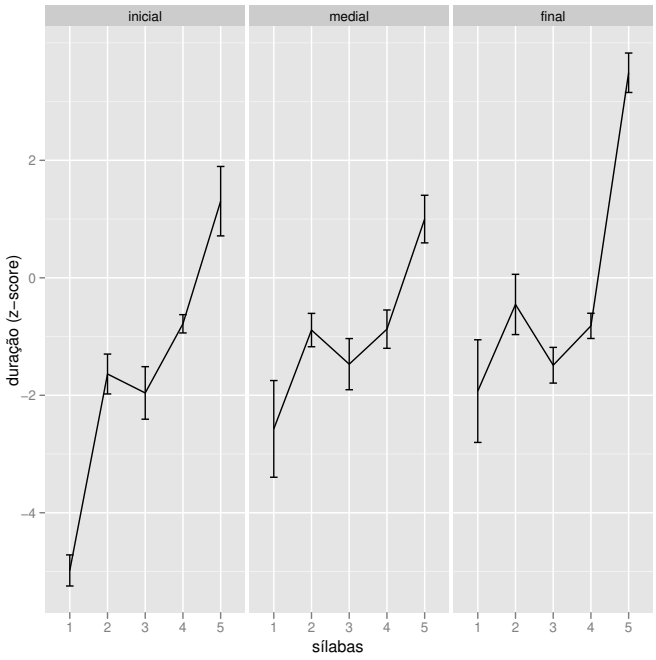


(b) Duração das sílabas das palavras do grupo P3

Figura 4.4



(a) Duração das unidades V-V das palavras do grupo O3



(b) Duração das sílabas das palavras do grupo O3

Figura 4.5

Grupo P4

A figura 4.6 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

- Unidades V-V

  Inicial $F(6,87) = 23\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,1$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 (n.s.)

  Medial $F(6,85) = 12,5\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,01$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,01$)

  Final $F(5,78) = 26,8\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,05$), 1 e 3 ($p < 0,001$), 1 e 4 ($p < 0,001$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 (n.s.), 3 e 5 ($p < 0,1$)

- Sílabas

  Inicial $F(6,91) = 41,6\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,001$), 3 e 4 (n.s.), 4 e 5 ($p < 0,01$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 ($p < 0,1$), 2 e 4 ($p < 0,01$), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,1$)

  Medial $F(6,93) = 6,5\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,01$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,01$), 3 e 5 (n.s.)

  Final $F(6,97) = 29,8\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,1$), 3 e 4 (n.s.), 4 e 5 ($p < 0,01$), 1 e 3 ($p < 0,05$), 1 e 4 ($p < 0,001$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 (n.s.)

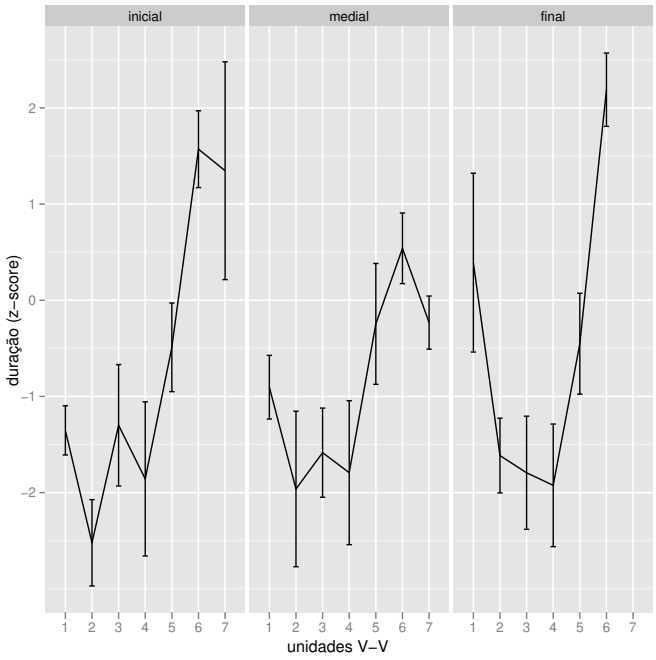
Grupo O4

A figura 4.7 mostra a média de duração das unidades V-V e sílabas em cada nível da variável sintática.

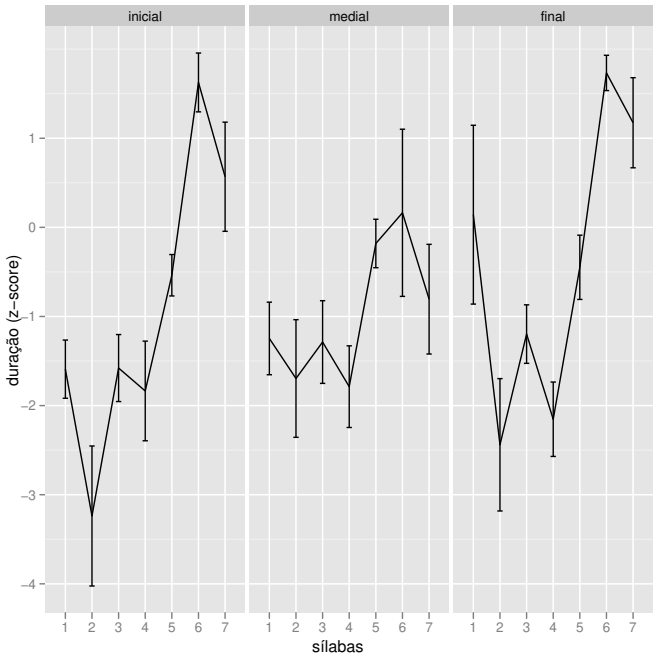
- Unidades V-V

  Inicial $F(5,78) = 23,3\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,05$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 ($p < 0,001$), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,001$)

  Medial $F(5,73) = 7,6\ p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 ($p < 0,01$), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,001$)



(a) Duração das unidades V-V das palavras do grupo P4



(b) Duração das sílabas das palavras do grupo P4

Figura 4.6

Final $F(4,67)12,5 =\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,001$), 1 e 3 ($p < 0,1$), 1 e 4 ($p < 0,01$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,001$)

- Sílabas

Inicial $F(5,80) = 26,6\ p < 0,001$; posições 1 e 2 ($p < 0,5$), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 ($p < 0,05$), 1 e 5 ($p < 0,001$), 2 e 4 (n.s.), 2 e 5 (n.s.), 3 e 5 ($p < 0,01$)

Medial $F(5,82) = 18,7\ p < 0,001$; posições 1 e 2 ($p < 0,05$), 2 e 3 ($p < 0,001$), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 ($p < 0,1$), 2 e 5 (n.s.), 3 e 5 (n.s.)

Final $F(5,79) = 34\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,001$), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 (n.s.), 3 e 5 (n.s.)

Discussão parcial

Os contornos de duração normalizada de unidades V-V e sílabas não mostram nenhum sinal claro de binarismo malgrado o tamanho da cadeia de pré-tônicas, posição sintática e tipo acentual. Em comparação com os resultados de duração do experimento de produção I, descritos na seção 3.2 nota-se aqui uma tendência bem menos acentuada à formação de proeminência acentual por meio do alongamento da unidade V-Vinicial, possivelmente em função do locutor, sugerindo que a proeminência inicial tem um caráter facultativo.

No grupo O4, a duração da segunda sílaba do grupo acentual (primeira da palavra morfológica) é mais alongada do que suas vizinhas, menos na posição inicial em que é apenas mais longa do que a sílaba anterior.

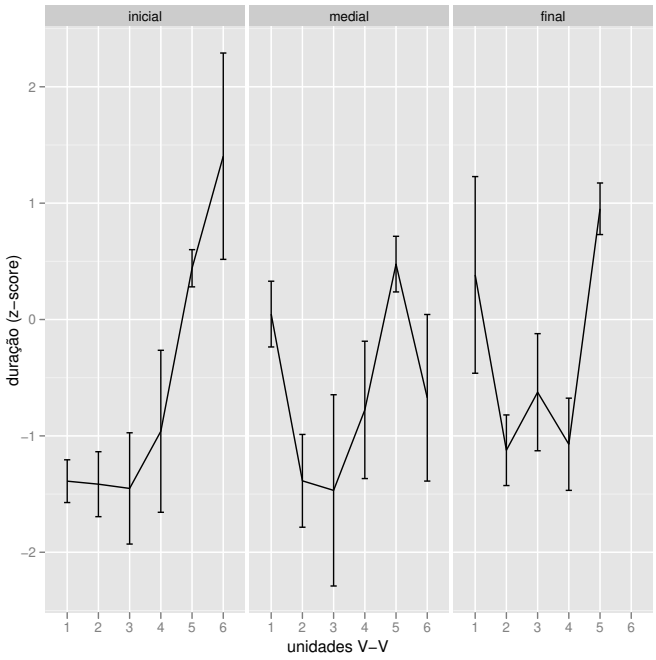
4.5.2 Efeito da variável discursiva

Nesta seção serão analisados os efeitos da variável discursiva sobre os contornos de duração de unidades V-V e sílabas em cada um dos seis grupos de palavras descritos no início da seção 4.5. Cada nível da variável discursiva (referente 'novo' ou 'velho') foi analisado separadamente. Para cada nível aplicaram-se duas ANOVAs de um fator tendo como variável independente as posições sintagmáticas e como variável dependente a duração normalizada das unidades V-V ou das sílabas.

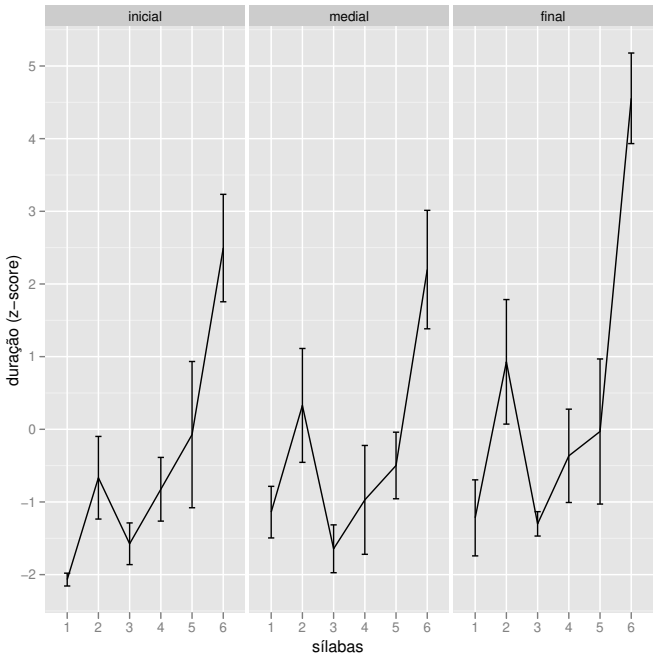
Grupo P2

A figura 4.8 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V



(a) Duração das unidades V-V das palavras do grupo O4



(b) Duração das sílabas das palavras do grupo O4

Figura 4.7

Novo $F(4,70) = 11\ p < 0,001$; posições 1 e 2 (n. s.), 2 e 3 (n. s.), 1 e 3 (n. s.)

Velho $F(4,70) = 10,7\ p <$; posições 1 e 2 (n. s.), 2 e 3 (n. s.), 1 e 3 (n. s.)

- Sílabas

  Novo $F(4,69) = 34,8\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n. s.), 1 e 3 ($p < 0,001$)

  Velho $F(4,70) = 51,1\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n. s.), 1 e 3 ($p < 0,001$)

Grupo O2

A figura 4.9 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V

  Novo $F(3,55) = 8,6\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,1$), 1 e 3 ($n.s.$)

  Velho $F(3,55) = 50,7\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,001$), 1 e 3 ($p < 0,001$)

- Sílabas

  Novo $F(3,56) = 4,6\ p < 0,01$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 1 e 3 (n.s.)

  Velho $F(3,56) = 96,3\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,001$), 1 e 3 ($p < 0,001$)

Grupo P3

A figura 4.10 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V

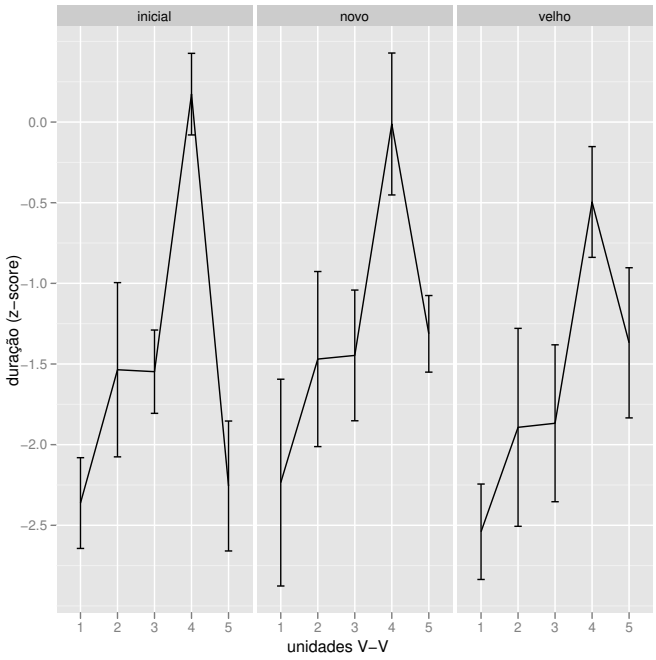
  Novo $F(5,79) = 10,6\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,01$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

  Velho $F(5,79) = 8,5\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

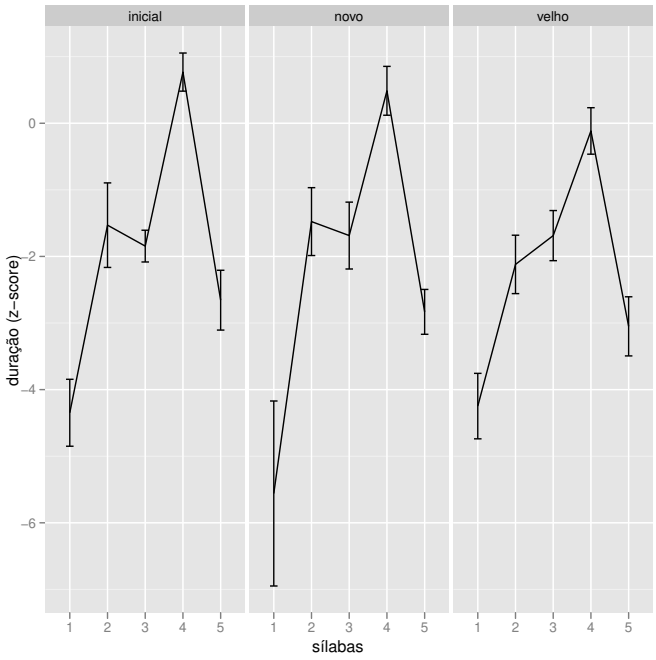
- Sílabas

  Novo $F(5,79) = 14,5\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 (n.s.)

  Velho $F(5,79) = 13,6\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 (n.s.)

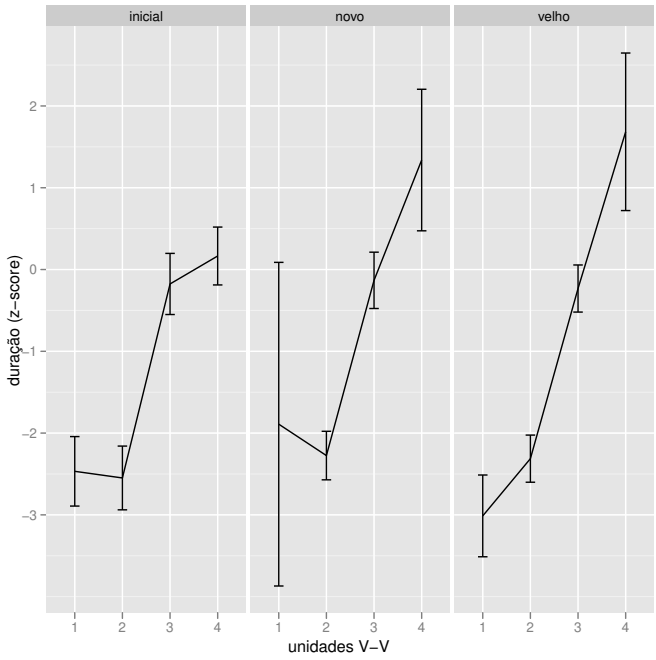


(a) Duração das unidades V-V das palavras do grupo P2

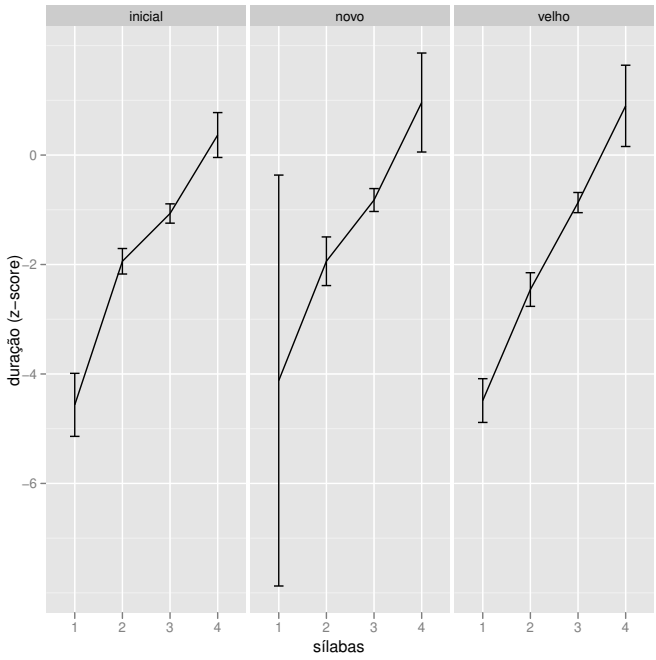


(b) Duração das sílabas das palavras do grupo P2

Figura 4.8

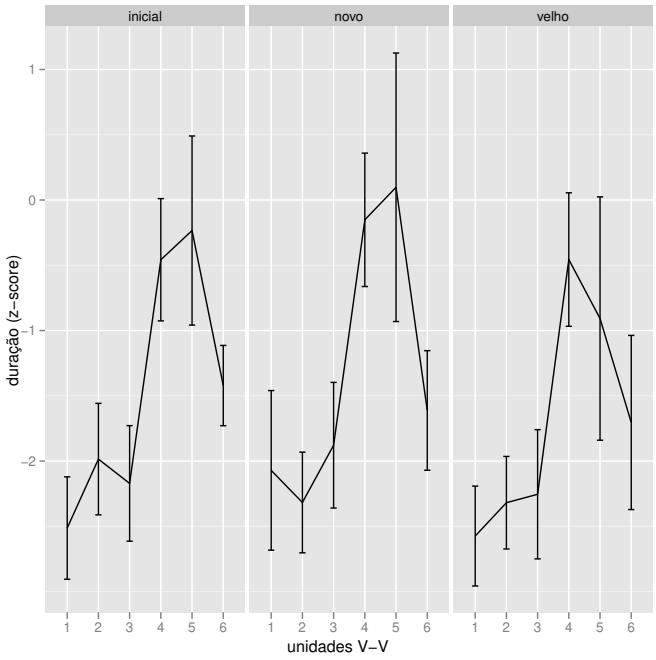


(a) Duração das unidades V-V das palavras do grupo O2

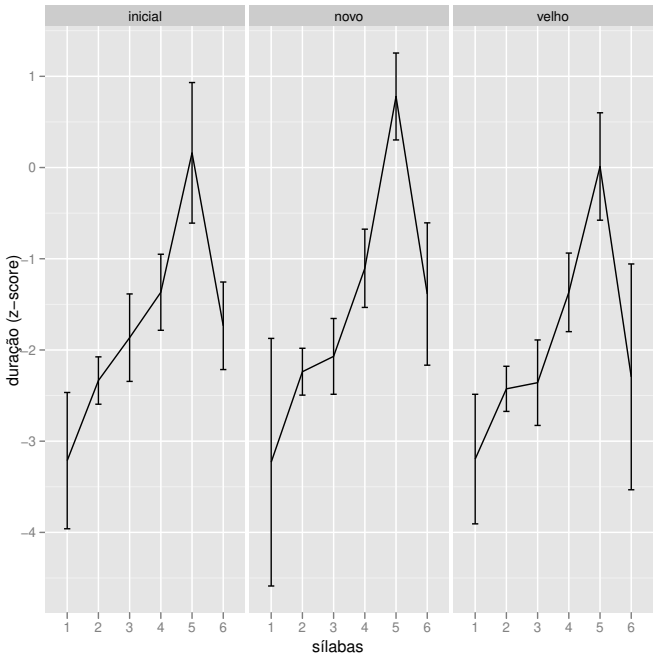


(b) Duração das sílabas das palavras do grupo O2

Figura 4.9



(a) Duração das unidades V-V das palavras do grupo P3



(b) Duração das sílabas das palavras do grupo P3

Figura 4.10

Grupo O3

A figura 4.11 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V

  Novo $F(4,68) = 81\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 ($p < 0,05$), 3 e 4 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,001$)

  Velho $F(4,68) = 66\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 ($p < 0,1$), 1 e 4 ($p < 0,01$), 2 e 4 ($p < 0,01$)

- Sílabas

  Novo $F(4,70) = 116,2\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 3 e 4 ($p < 0,05$), 1 e 3 ($p < 0,001$), 1 e 4 ($p < 0,001$), 2 e 4 (n.s.)

  Velho $F(4,70) = 110,5\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 (n.s.), 3 e 4 ($p < 0,001$), 1 e 3 ($p < 0,001$), 1 e 4 ($p < 0,001$), 2 e 4 ($p < 0,05$)

Grupo P4

A figura 4.12 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V

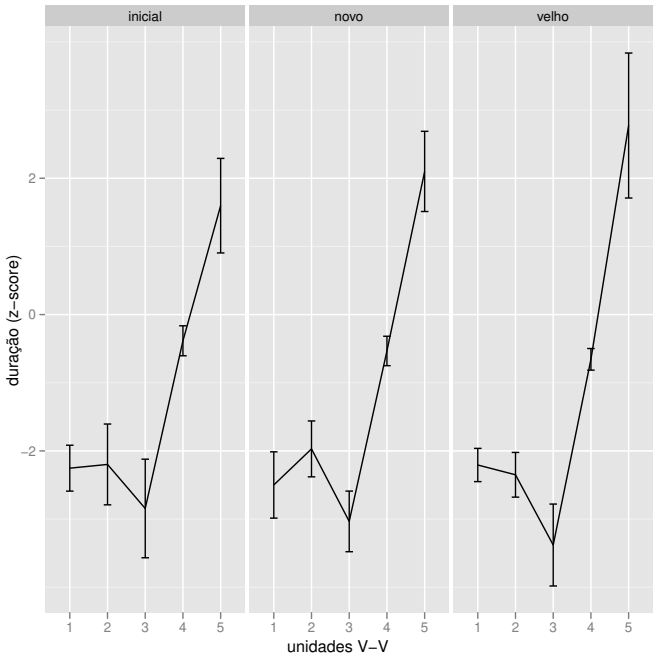
  Novo $F(6,95) = 23,3\ p < 0,001$; posições 1 e 2 ($p < 0,05$), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,001$), 1 e 3 ($p < 0,1$), 1 e 4 ($p < 0,001$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,01$)

  Velho $F(6,94) = 28,4\ p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,1$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,05$)

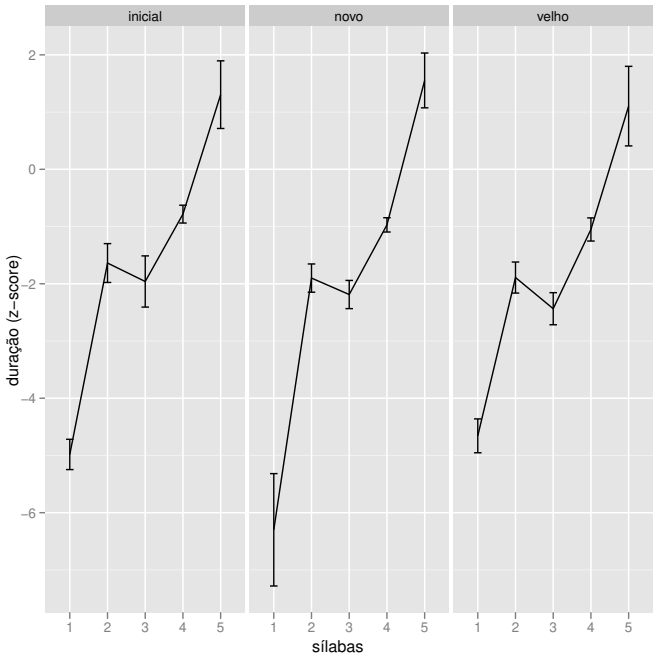
- Sílabas

  Novo $F(6,98) = 29\ p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 ($p < 0,5$), 3 e 4 ($p < 0,1$), 4 e 5 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 ($p < 0,5$), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 (n.s.)

  Velho $F(6,98) = 49,6\ p < 0,001$; posições 1 e 2 ($p < 0,001$), 2 e 3 ($p < 0,001$), 3 e 4 (n.s.), 4 e 5 ($p < 0,001$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 ($p < 0,01$), 2 e 4 ($p < 0,01$), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,01$)

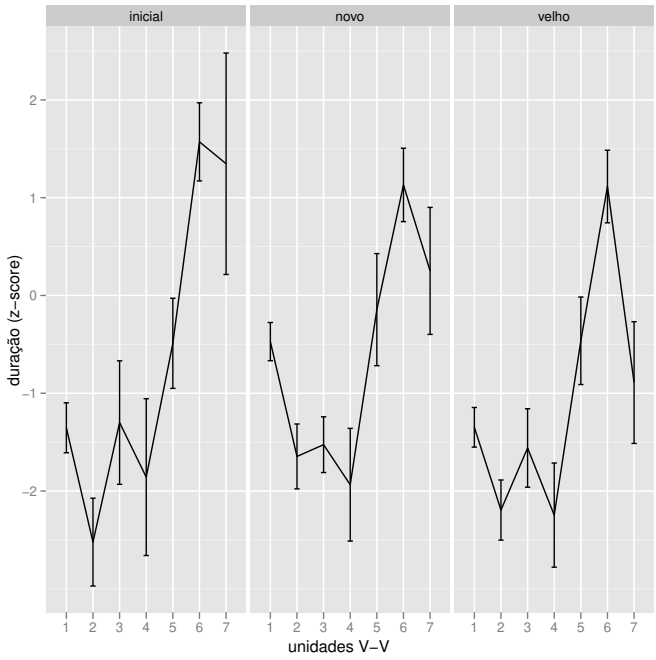


(a) Duração das unidades V-V das palavras do grupo O3

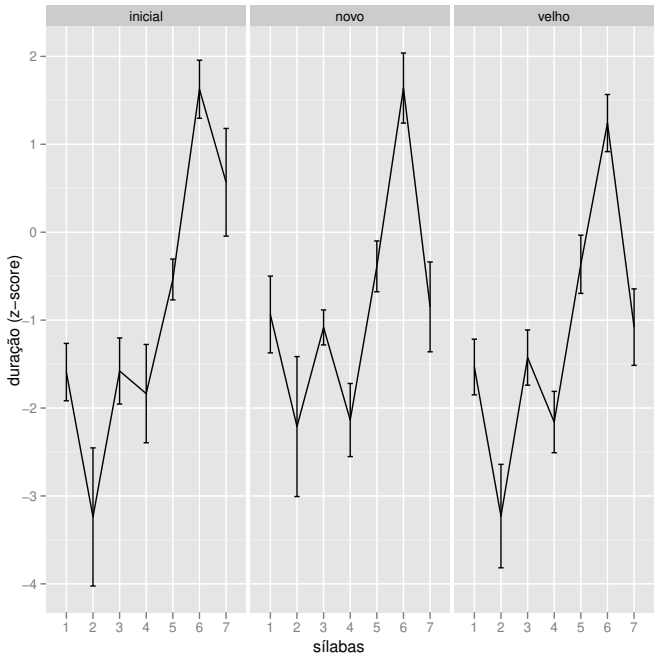


(b) Duração das sílabas das palavras do grupo O3

Figura 4.11



(a) Duração das unidades V-V das palavras do grupo P4



(b) Duração das sílabas das palavras do grupo P4

Figura 4.12

Grupo O4

A figura 4.13 mostra a média de duração das unidades V-V e sílabas nos dois níveis da variável discursiva. Para fins de comparação mostra-se também o contorno da condição 'inicial' da variável sintática.

- Unidades V-V

  Novo $F(5,80) = 7$ $p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 ($p < 0,01$), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,01$), 3 e 5 (n.s.)

  Velho $F(5,75) = 5,4$ $p < 0,001$; posições 1 e 2 (n.s.), 2 e 3 (n.s.), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 ($p < 0,1$), 2 e 4 (n.s.), 2 e 5 ($p < 0,001$), 3 e 5 ($p < 0,01$)

- Sílabas

  Novo $F(5,83) = 23,7$ $p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 ($p < 0,001$), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 ($p < 0,5$), 3 e 5 (n.s.)

  Velho $F(5,82) = 18,2$ $p < 0,001$; posições 1 e 2 ($p < 0,01$), 2 e 3 ($p < 0,05$), 3 e 4 (n.s.), 4 e 5 (n.s.), 1 e 3 (n.s.), 1 e 4 (n.s.), 1 e 5 (n.s.), 2 e 4 (n.s.), 2 e 5 (n.s.), 3 e 5 (n.s.)

Duração da palavra

Os resultados de Fowler e Housum (1987) citados na seção 4.3 sugerem que se compare a duração total da palavra-alvo para verificar possíveis efeitos da variável pragmático-discursiva. Nesse caso faz sentido comparar a duração bruta da palavra inteira (descontando o artigo, já que eles não são os mesmos nas duas condições) uma vez que comparadas aos pares a única diferença entre elas é o *status* referencial.

A tabela 4.3 apresenta a média de duração em milissegundos das palavras de cada grupo nas duas condições (referentes novo e dado). As médias foram comparadas por meio de testes-t pareados com os graus de liberdade corrigidos em caso de heterocedasticidade das amostras[6].

nº pret.	acento	novo	dado	p
2	parox.	575	532	0,028
	oxít.	499	485	n.s.
3	parox.	770	682	0,004
	oxít.	652	633	n.s.

[6]Testes de Levene comparando a variância das amostras de status 'novo' e 'velho' em cada grupo de palavras indicaram apenas um caso de heterocedasticidade (grupo O4).

4	parox.	968	884	0,014
	oxít.	980	851	$< 0,001$

Tabela 4.3: Duração média (em milissegundos) das palavra-chave de diferentes tamanhos e tipos acentuais nas duas condições da variável pragmático-discursiva. A quinta coluna mostra o valor de p das comparações pareadas entre as duas condições.

Discussão parcial

A comparação entre os padrões de duração das palavras dos grupos P2, O2, P3 e O3 mostra que os contornos duracionais tanto das unidades V-V quanto das sílabas é sempre ascendente ou estacionário a despeito do *status* referencial. Não se observa um padrão que pudesse ter sido gerado por um mecanismo de alternância binária ou mesmo um padrão de proeminência inicial.

No grupo P4, a duração das unidades V-V indica haver alongamento inicial, mas apenas na condição de referente novo. No grupo O4 a alongamento ocorre nas duas condições. A duração das sílabas no grupo P4 revela um alongamento da sílaba inicial nas duas condições de status referencial e no grupo 04 é a segunda sílaba que exibe o alongamento relativo.

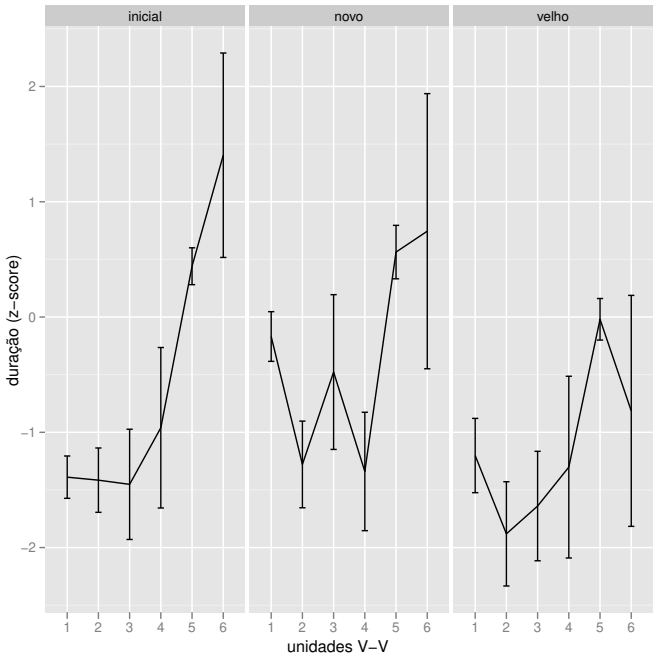
Os dados de duração bruta da palavra-chave mostram que tende a haver um encurtamento da duração total da palavra quando ela é um referente velho ou dado no discurso mesmo na fala lida. Essa tendência não é condicionada pelo número de sílabas da palavra, embora não seja universal no *corpus* uma vez que as oxítonas parecem menos propensas ao encurtamento do que as paroxítonas.

Como o contraste entre os contornos duracionais de palavras 'novas' e 'velhas' não mostram diferenças marcantes, isto é, a evolução da duração ao longo da palavra parece seguir a mesma dinâmica nos dois casos, então o encurtamento das palavras-chave na condição 'velho' não poderia ser explicado pela ausência, em comparação com palavras na condição 'novo', de alongamentos significativos de alguma sílaba específica. O encurtamento parece ser melhor acomodado como o resultado de uma pronúncia hipoarticulada da palavra como um todo nas segundas ocorrências e não pela perda mais ou menos pontual de uma proeminência.

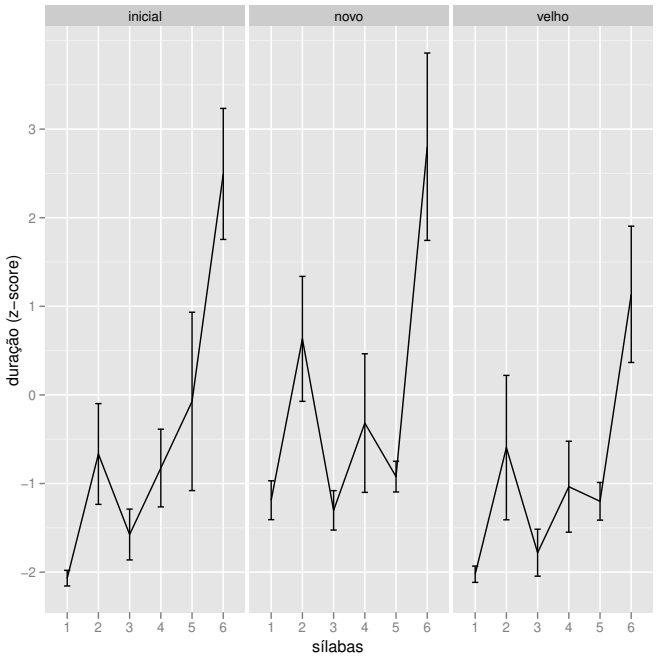
4.6 Resultados: frequência fundamental

A influência das variáveis independentes do *corpus* sobre a f_0 será avaliada sob duas perspectivas: os efeitos globais e a distribuição de pontos extremos extraídos por meio do algoritmo descrito no apêndice B.

A variação global da f_0 é caracterizada por meio de descritores simples de tendência central e dispersão: média (μ), desvio-padrão (σ), gama ou amplitude de variação (diferença entre o maior e o menor valor da amostra), valor máximo e valor mínimo. Os valores foram calculados em Hertz (Hz) e semitons re 100Hz (st) sobre a amostra de f_0 no intervalo de unidades V-V contendo a palavra-chave em cada frase-veículo do *corpus*.



(a) Duração das unidades V-V das palavras do grupo O4



(b) Duração das sílabas das palavras do grupo O4

Figura 4.13

4.6.1 Tamanho da palavra-chave

Nesta seção serão investigados os efeitos da variável lexical tamanho da palavra sobre a variação global da f_0. A tabela 4.4 mostra os valores médios dos descritores estatísticos em cada um dos 3 grupos de palavras de tamanhos diferentes (palavras com 2, 3 ou 4 sílabas pré-tônicas). Palavras oxítonas e paroxítonas foram tomadas em conjunto uma vez que o interesse é a caracterização da variação global da f_0 na palavra-chave. Foram aplicados testes de ANOVA tendo o tamanho da palavra como variável independente e o descritor estatístico de cada linha como variável dependente para verificar se as diferenças entre os valores dos descritores de cada grupo de palavras é significativa. O valor reportado na quinta coluna é o valor de p de cada ANOVA.

variável	2 pret.	3 pret.	4 pret.	p
σ (Hz)	11,3	14	17,7	$< 0,001$
σ (st)	1,4	1,7	2,1	$< 0,001$
gama (Hz)	43,2	52,2	65,5	$< 0,001$
gama (st)	5,2	6,4	7,7	$< 0,001$
máx. (Hz)	165	167	178	$< 0,001$
máx. (st)	8,6	8,8	9,8	$< 0,001$
mín. (Hz)	122	115	113	$< 0,001$
mín. (st)	3,4	2,4	2	$< 0,001$

Tabela 4.4: Médias dos parâmetros da variação global de f_0 separadas por tamanho da palavra.

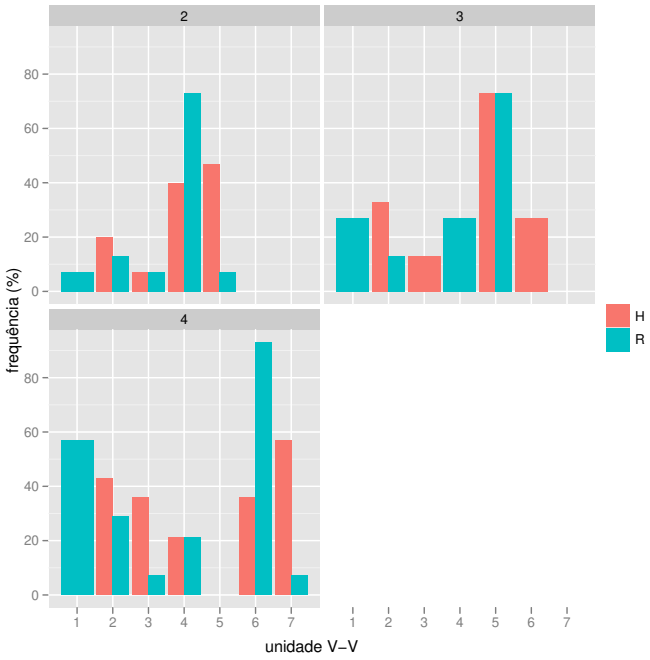
A tabela mostra que há um claro efeito do aumento da dispersão na distribuição de f_0 conforme a palavra-chave tem seu tamanho aumentado pelo acréscimo de sílabas pré-tônicas. Observa-se também uma tendência a valores mais extremos com o aumento do tamanho da palavra, como evidenciam os valores da gama de variação e dos máximos e mínimos.

4.6.2 Variável sintática

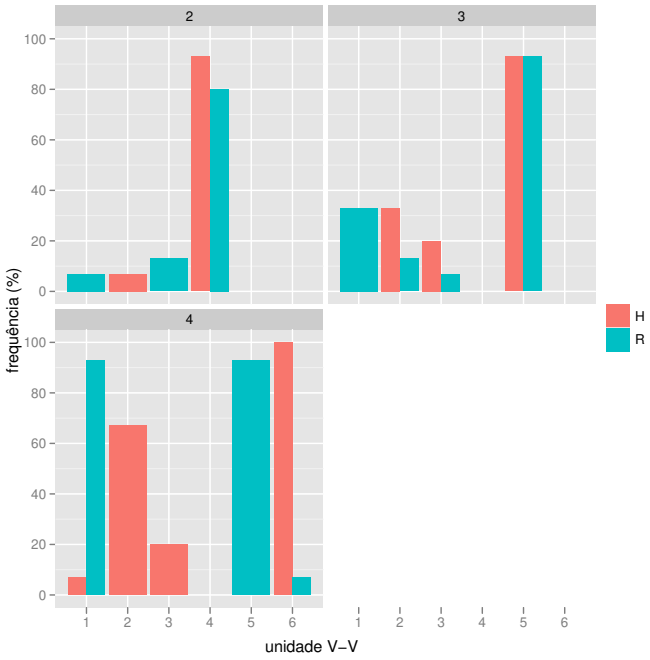
Esta seção relata dados sobre a distribuição de extremos de f_0 altos (H) e de f_0' ascendentes (R) (de acordo com o algoritmo apresentado no apêndice B) ao longo das palavras-chave e a influência da variável sintática sobre essa distribuição. As tabelas apresentadas nas seções a seguir quantificam a variabilidade na ocorrência de pontos extremos de f_0 nos diferentes grupos de palavras do *corpus*.

Distribuição de pontos extremos H: posição inicial

A tabela 4.5 mostra a porcentagem de ocorrências de diferentes configurações tonais (contornos com diferentes números de pontos H) nos grupos de palavras-alvo. A figura 4.14 mostra a frequência relativa dos extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.



(a) Palavras paroxítonas



(b) Palavras oxítonas

Figura 4.14: Posição sintática inicial — frequência relativa de extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

nº pret.	2	3	4
parox.	1 H: 80% 2 Hs: 20%	1 H: 27% 2 Hs: 73%	1 H: 7% 2 Hs: 79% 3 Hs: 14%
oxít.	1 H: 67% 2 Hs: 33%	1 H: 33% 2 Hs: 67%	2 Hs : 100%

Tabela 4.5: Porcentagem de ocorrências de pontos extremos H por grupo de palavras em posição sintática inicial.

Distribuição de pontos extremos H: posição medial

A tabela 4.6 mostra a porcentagem de ocorrências de diferentes configurações tonais nos grupos de palavras-alvo. A figura 4.15 mostra a frequência relativa dos extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

nº pret.	2	3	4
parox.	1 H: 77% 2 Hs: 23%	1 H: 86% 2 Hs: 14%	1 H: 54% 2 Hs: 46%
oxít.	1 H: 93% 2 Hs: 7%	1 H: 70% 2 Hs: 30%	1 H: 36% 2 Hs: 64%

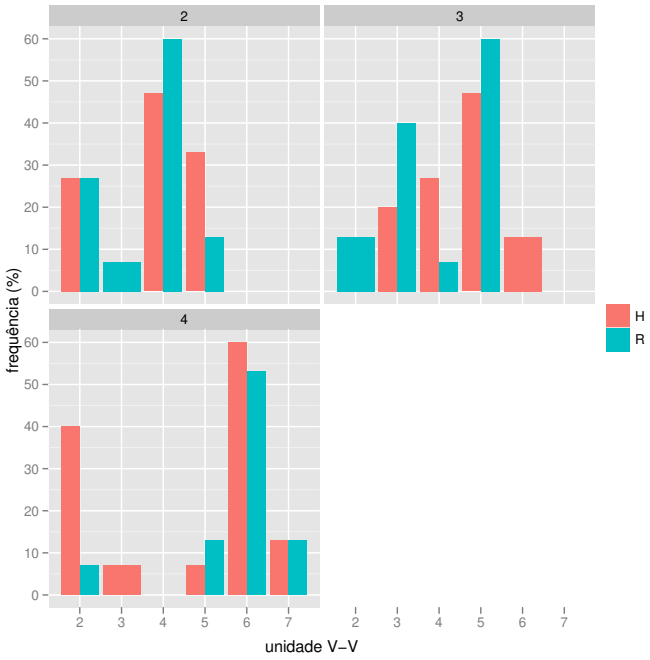
Tabela 4.6: Porcentagem de ocorrências de pontos extremos H por grupo de palavras em posição sintática medial.

Distribuição de pontos extremos H: posição final

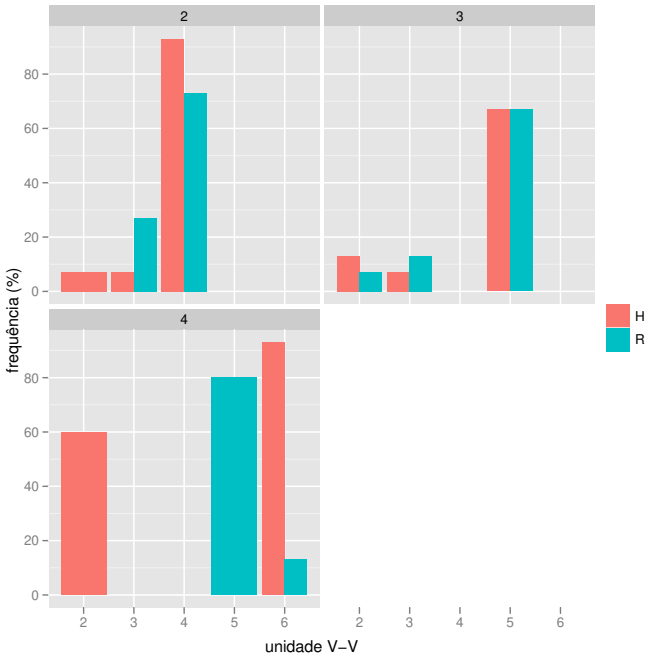
A tabela 4.7 mostra a porcentagem de ocorrências de diferentes configurações tonais nos grupos de palavras-alvo.

nº pret.	2	3	4
parox.	1 H: 79% 2 Hs: 21%	1 H: 67% 2 Hs: 33%	1 H: 93% 2 Hs: 7%
oxít.	1 H: 93% 2 Hs: 7%	1 H: 40% 2 Hs: 60%	1 H: 60% 2 Hs: 40%

Tabela 4.7: Porcentagem de ocorrências de pontos extremos L por grupo de palavras em posição sintática final.



(a) Palavras paroxítonas



(b) Palavras oxítonas

Figura 4.15: Posição sintática medial — frequência relativa de extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

Discussão parcial

Quanto à configuração tonal, o aumento do número de sílabas pré-tônicas tem o efeito de aumentar a ocorrência de configurações bitonais, isto é, configurações em que ocorrem dois extremos de f_0 do tipo H. Há uma interação, no entanto, entre aumento de ocorrências configurações com mais de um extremo H e a variável sintática: o aumento das configurações bitonais é maior quando as palavras-chave estão em posição sintática inicial. As configurações com três Hs foram registradas — apenas 5 num universo de aproximadamente 360 repetições (cf. tabelas 4.5 e 4.9) — ocorrem em posição sintática inicial (sintaticamente, os referentes novos também estão em posição inicial). Pode-se observar, na distribuição das configurações tonais, um efeito do tipo acentual das palavras-chave. A menos de algumas exceções, com o aumento das sílabas pré-tônicas as oxítonas tem mais chance de apresentarem configurações bitonais do que as paroxítonas.

A influência do tamanho da cadeia de pré-tônicas sobre a distribuição dos extremos de f_0 H e de f_0' R ao longo das unidades V-V manifesta-se na forma de um aumento na frequência de extremos H e R associados às unidades V-V iniciais, malgrado a posição sintática. Não é possível distinguir um efeito claro da posição sintática e do tipo acentual sobre o alinhamento dos extremos em relação às unidades V-V.

Principalmente a partir dos dados das palavras com 4 sílabas pré-tônicas é possível descartar a manifestação de um binarismo estrito na atribuição de extremos H. Em palavras desse tipo, como *possibilidade* e *peregrinação*, a hipótese do binarismo prevê acentos secundários nas sílabas "po" e "bi" na primeira palavra e nas sílabas "pe" "gri" na segunda. Portanto, caso a hipótese fosse verdadeira esperar-se-ia um grande número de configurações tritonais (dois extremos H para os acentos secundários e um para a tônica), o que definitivamente não é o caso. A melhor interpretação para os dados parece ser a da existência de um tom alto (resultante do extremo H) associado a uma das pré-tônicas, em geral a primeira ou a segunda da palavra morfológica. Com o aumento da cadeia de pré-tônicas esse extremo pode se associar a qualquer pré-tônica, com exceção da imediatamente anterior à tônica.

4.6.3 Efeito da variável discursiva

A seção 4.6.3 apresenta o efeito da variável discursiva sobre o padrão global de variação de f_0 e as seções seguintes relatam dados sobre a distribuição de pontos extremos H ao longo das palavras-chave e a influência da variável discursiva sobre essa distribuição.

Efeito global

Nesta seção serão investigados os efeitos da variável discursiva sobre a sobre a variação global da f_0. A tabela 4.8 mostra os valores médios dos descritores estatísticos nos dois níveis da variável discursiva. Para o cálculo da média os dados não foram separados por tamanho da palavra ou tipo acentual. O intervalo tomado para análise é a palavra-chave e o artigo que a precede. A comparação entre as médias foi realizada por meio de testes-t de amostras

independentes. O valor reportado na quarta coluna é o valor de p de cada teste. Em nenhum dos testes foi diagnosticada heterogeneidade de variância entre as amostras comparadas.

variável	novo	dado	p
μ (Hz)	151	132	$< 0,001$
μ (st)	7	4,7	$< 0,001$
σ (Hz)	17	11,7	$< 0,001$
σ (st)	2	1,5	$< 0,001$
gama (Hz)	63,8	43,5	$< 0,001$
gama (st)	7,3	5,6	$< 0,001$
máx. (Hz)	184	156	$< 0,001$
máx. (st)	10,5	7,7	$< 0,001$
mín. (Hz)	121	112	$< 0,001$
mín. (st)	3,2	2	$< 0,001$

Tabela 4.8: Médias dos parâmetros da variação global de f_0 separadas por *status* referencial .

A inspeção da tabela mostra que a manipulação do *status* referencial da palavra-chave causa efeitos na dispersão na distribuição de f_0. Os descritores de dispersão das palavras-chave que são referentes novos tem valores sistematicamente maiores do que as mesmas palavras quando são referentes velhos ou dados. Observa-se, ainda, uma tendência a valores mais extremos com o aumento da tamanho da palavra, como evidenciam os valores da gama de variação e dos máximos e mínimos.

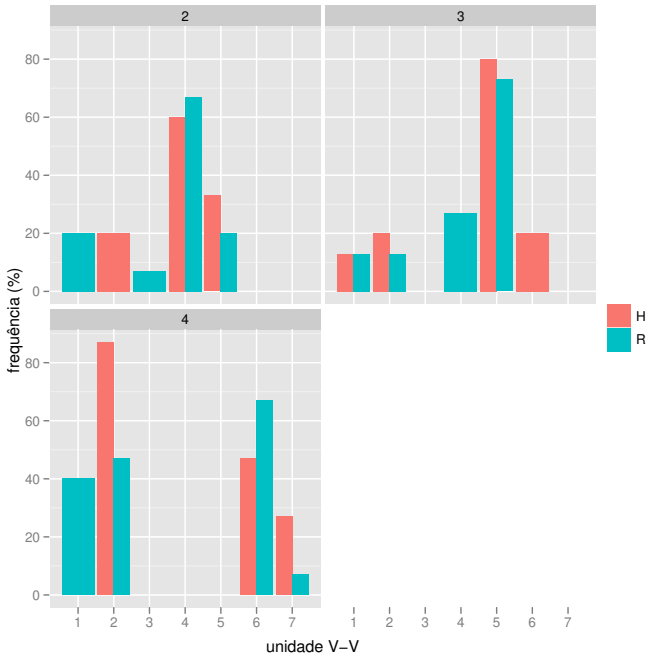
É interessante comparar os presentes resultados com os apresentados na tabela 4.4, que mostram um consistente relação entre aumento na variabilidade global de f_0 motivada pelo aumento no tamanho de palavra. Como a variável tamanho da palavra e a variável discursiva variam de forma ortogonal no *corpus* isso garante que os dois efeitos são independentes.

Distribuição de pontos extremos H: referente novo

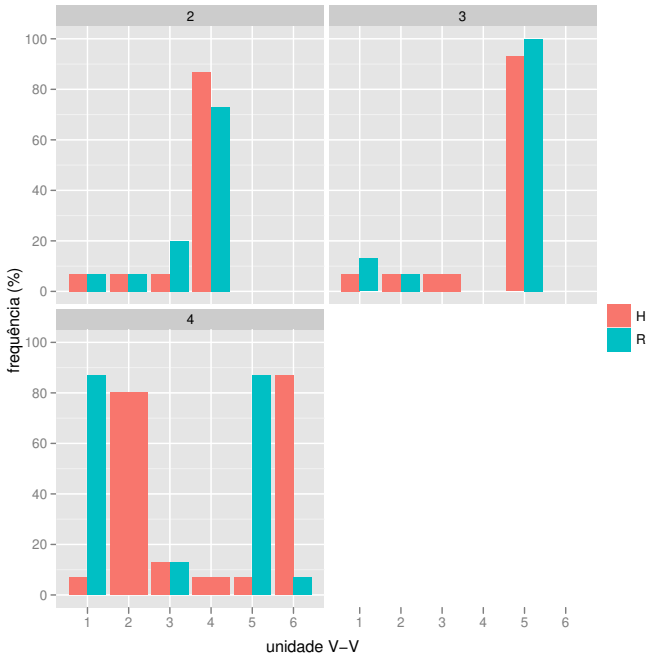
A tabela 4.9 mostra a porcentagem de ocorrências de diferentes configurações tonais nos grupos de palavras-alvo. A figura 4.16 mostra a frequência relativa dos extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

nº pret.	2	3	4
parox.	1 H: 87% 2 Hs: 13%	1 H: 67% 2 Hs: 33%	1 H: 29% 2 Hs: 71%
oxít.	1 H: 93% 2 Hs: 7%	1 H: 79% 2 Hs: 21%	1 H: 13% 2 Hs: 73% 3 Hs: 13%

Tabela 4.9: Porcentagem de ocorrências de pontos extremos H por grupo de palavras em condição de referente novo.



(a) Palavras paroxítonas



(b) Palavras oxítonas

Figura 4.16: Referentes novos — frequência relativa de extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

Distribuição de pontos extremos H: referente velho

A tabela 4.10 mostra a porcentagem de ocorrências de diferentes configurações tonais nos grupos de palavras-alvo. A figura 4.17 mostra a frequência relativa dos extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

nº pret.	2	3	4
parox.	1: 100%	1: 73%	1: 36%
		2: 27%	2: 64%
oxít.	1: 92%	1: 67%	1: 7%
	2: 8%	2: 33%	2: 93%

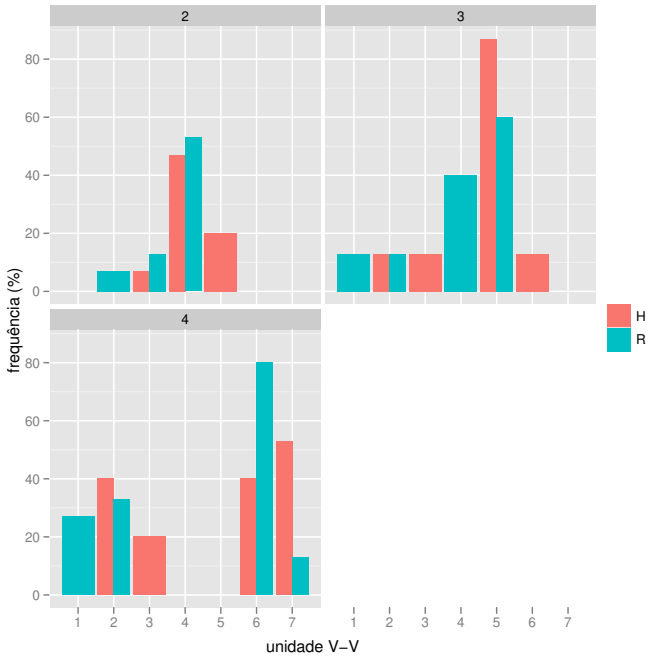
Tabela 4.10: Porcentagem de ocorrências de pontos extremos H por grupo de palavras em condição de referente velho.

Discussão parcial

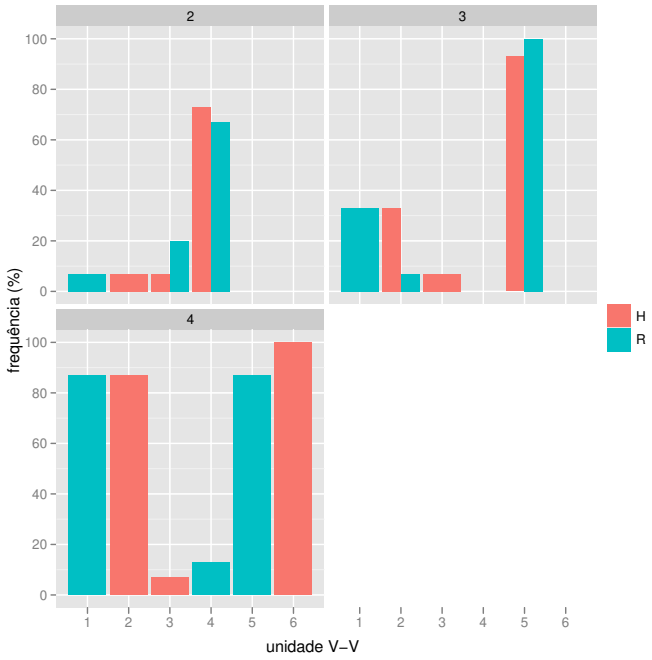
O contraste entre os dois níveis da variável pragmático-discursiva não parece exercer nenhum efeito expressivo sobre a distribuição de configurações com diferentes números de extremos do tipo H. As porcentagens de configurações com 1 e 2 extremos H são bastante semelhantes quando comparadas palavras de mesmo tipo acentual e número de sílabas pré-tônicas. Observa-se a existência de um efeito de tamanho da cadeia pré-tônica no sentido de aumentar o número de configurações com mais um extremo H à medida que o número de pré-tônicas aumenta. Não se depreende dos dados um efeito claro devido ao tipo acentual.

Na distribuição de extremos H e R ao longo das unidades V-V pode-se observar o aumento da frequencia destes extremos associados à unidades V-V iniciais à medida que aumenta o número de pré-tônicas. Em relação ao tipo acentual, a diferença mais expressiva é a maior frequência de extremos H iniciais nas palavras oxítonas quando o status referencial das palavras é de referente velho, malgrado o número de pré-tônicas. O efeito da manipulação do *status* referencial interage com o tipo acentual da palavra-chave. As paroxítonas que são também referentes novos, principalmente os grupos P2 e P4, apresentam maior frequência de extemos H associados a unidades V-V iniciais quando comparadas com os grupos correpondente com *status* de referente velho. No grupo das oxítonas com três pré-tônicas (grupo O3), o efeito do status referencial é contraditório com a expectativa, pois as palavras em condição de referente velho apresentam maior frequência de ocorrência de extremos H associado à unidade V-V 2.

Os efeitos mais expressivos da variável pragmático-discursiva parecem estar concentrados sobretudo na variação global da curva de f_0, descrita na seção 4.6.3 e sumarizada na tabela 4.8, e em menor grau em diferenças na quantidade e alinhamento de extremos do tipo H. A manipulação do *status* referencial, portanto, pelo menos no *corpus* do presente experimento, não ocasionou o aparecimento de proeminências secundárias que pudessem diferenciar a condição



(a) Palavras paroxítonas



(b) Palavras oxítonas

Figura 4.17: Referentes velhos — frequência relativa de extremos do tipo H (alto) e R (ascendente) por unidade V-V nos grupos de palavras com 2, 3 e 4 pré-tônicas.

considerada prosodicamente mais forte (referente novo) da mais fraca (referente novo).

4.7 Simulação com o modelo de ritmo

Nesta seção são apresentados os resultados de uma simulação realizada com o modelo dinâmico do ritmo. O objetivo da simulação é investigar a capacidade do modelo de reproduzir os aspectos dos dados empíricos de duração obtidos nos experimentos de produção I e II desta tese que importam para a análise da acentuação secundária. Os dados empíricos não apontam a ocorrência sistemática de padrões de atribuição de acento frasal que sejam compatíveis com a hipótese de binaridade estrita. Espera-se, portanto, que o modelo seja capaz de gerar contornos de duração com as características observadas nos contornos naturais. Em especial, espera-se que as simulações alcancem os seguintes objetivos:

1. o modelo deve ser capaz de reproduzir a proeminência inicial manifesta como um alongamento da primeira unidade V-V do grupo acentual atestada nos dados naturais;

2. em grupos acentuais com maior número de unidades V-V deve haver mais chance de ocorrer alongamento inicial;

3. a ocorrência do alongamento inicial não deve ser obrigatória, uma vez que comparando os resultados do experimento I e II desta tese é possível dizer que a ocorrência proeminência inicial na forma do alongamento da primeira unidade V-V do grupo acentual ocorreu de forma mais sistemática no primeiro experimento do que no segundo.

4.7.1 Procedimento

Para realizar a simulação foi usado o *script* de *MatLab* (The Mathworks 2001) `ritmar`, cujo código-fonte é apresentado no apêndice A.5 de Barbosa (2006, p. 480–485). O programa é uma implementação computacional da modelo dinâmico do ritmo. Na versão do *script* Ritmar utilizada o usuário deve especificar valores para os parâmetros independentes do modelo[7], α, β, w_0, T_0 e i, além de especificar quantos grupos acentuais quer simular e o número de unidades V-V em cada um deles. O resultado produzido pelo *script* é a duração abstrata (i.e., sem as perturbações impostas pela interação com a pauta gestual) das unidades V-V ao longo dos grupos acentuais, formando o que se poderia chamar de máscara ou contorno acentual. Em tese, esse contorno deve ser semelhante aos contornos duracionais naturais depois de normalizados e suavizados[8].

Para realizar a simulação foi necessário estipular valores para as variáveis independentes do modelo. Não adotou-se um procedimento tão completo como o descrito por Barbosa (2002) para a otimização dos parâmetros do modelo. O parâmetro T_0, valor aproximado (em segundos) do período do oscilador silábico ainda não induzido, foi o único para o qual realizou-se uma estimação a partir dos dados naturais. Para tanto, calculou-se a média conjunta de duração da primeira e da segunda unidades V-V do grupo de palavras de 4 sílabas do *corpus* do experimento

[7] Cf. seção 1.2 para o significa de cada parâmetro.

[8] Cf. apêndice A para uma descrição detalhada destes algoritmos de análise da duração.

de produção I, por ser o grupo com maior número de dados (2195 dados para as duas unidades referidas). O resultado foi $T_0 = 0,158$ segundos. Para os demais parâmetros, cuja otimização exigiria um trabalho considerável, foram utilizados os seguintes valores sugeridos em Barbosa (2002): $\alpha = 0,5$, $\beta = 0,8$, $w_0 = 0,7$, $T_0 = 0,158s$ e $i = 1$. O resultado das simulações pode ser observado na figura 4.18.

4.7.2 Grupos acentuais simulados

Foram simulados quatro grupos acentuais com o tamanho em unidades V-V variando entre cinco e oito, tamanhos similares aos encontrados nos *corpora* dos experimentos de produção I e II da tese. Alguns exemplos de estruturação rítmica com a consequente formação de agrupamentos acentuais de tamanho diferentes são mostrados abaixo.

(a) |ap at ar <u>at</u> ap| (5 V-Vs com culminância no V-V_3)
Ex. "|A patarata p|arece menor hoje"

(b) |ap at ar at ab ud <u>ist</u> ap| (8 V-Vs com culminância no V-V_6)
Ex. "|A patarata budista p|arece menor"

(c) |ad ir ig ib il id <u>ad</u> ep| (8 V-Vs com culminância no V-V_6)
Ex: "|A dirigibilidade p|arece menor hoje"

O contorno de duração produzido pelo modelo para um grupo acentual com oito unidades V-V deve ser o mesmo não importando se ele compreende duas palavras (como em **c**) ou três (como em **b**).

4.7.3 Resultados

A figura 4.18 mostra os contornos duracionais dos grupos acentuais simulados.

Pode-se considerar que a simulação alcança o objetivo 1 listado na introdução do capítulo, pois, há uma semelhança entre a forma geral do contorno dos grupos simulados e a forma do contorno normalizado dos dados comportamentais. Em relação ao objetivo 2, os resultados da simulação realizada nesta tese converge com os alcançados nas simulações reportadas na seção 6.5.1 de Barbosa (2006) que mostram que, conforme aumenta o número de unidades V-V em um grupo acentual (foram feitas simulações de grupos com 3, 6 e 8 unidades V-V), cresce o número de combinações dos parâmetros α e β que geram alongamento inicial (cf. figuras 6.22, 6.23 e 6.24, pp. 410–412 em Barbosa 2006). Não foi feita uma simulação em Barbosa (2006) para verificar se o aumento do grupo acentual gera alongamentos iniciais de maior magnitude como o exame da figura 4.18 nesta seção parece sugerir.

Em relação ao objetivo 3 a simulação realizada nesta tese não resolve definitivamente a questão, já que dos quatro grupos acentuais simulados três (os mais longos) apresentam alongamento inicial. As simulações de Barbosa (2006) citadas no parágrafo anterior, no entanto, especialmente o exame das figuras, mostra que mesmo para grupos acentuais maiores há faixas de combinação dos parâmetros α e β que não geral alongamento inicial.

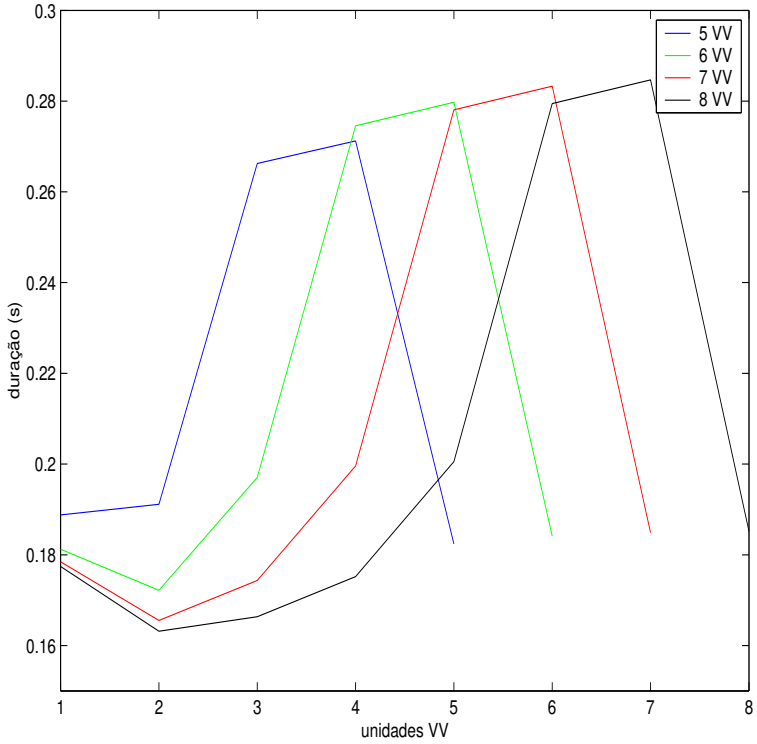


Figura 4.18: Contornos duracionais simulados para grupos acentuais com tamanho entre 5 e 8 unidades V-V.

Capítulo 5

Experimentos de percepção

Bela Julesz: So tell me, Georg, what do you think happiness is? Georg von Békésy: That is easy Bela. Happiness is a good experiment.

Bela Julesz, *Dialogues on perception*

Como foi dito na Introdução, um dos objetivos da tese é delinear uma proposta de abordagem, fundada na teoria dos sistemas dinâmicos, que trate de maneira integrada tanto a produção quanto a percepção do ritmo da fala. A seção 1.4 sugere que os mesmos mecanismos, a indução (*entrainment*) e acoplamento entre osciladores, regem o funcionamento dos dois diferentes aspectos do fenômeno do ritmo. Também na seção 1.4 sugere-se um esquema geral para essa abordagem unificada, baseado nos modelos de McAuley (1995) e Barbosa (2006). Um dos postulados importantes que a proposta apresentada na referida seção assume é que o modelo interno de ritmo dos falantes/ouvintes funciona de tal modo que, quando se está na posição de ouvinte, a sucessão de *onsets* vocálicos no sinal acústico a ser processado funciona como estímulo externo que induz o acoplamento do período do oscilador silábico do ouvinte com este estímulo externo, de forma que o período de suas oscilações seja uma espécie de espelho dos períodos do oscilador silábico induzido subjacente à produção do sinal acústico sendo processado. Para que essa postulação tenha validade empírica é preciso prover evidência de que os ouvintes usam a unidade V-V (entendida como o intervalo entre dois *onsets* vocálicos consecutivos) como unidade privilegiada de processamento dos contornos duracionais. Esse é um outro aspecto em que a produção e percepção podem ser vistas como atividades especulares, pois no modelo de produção do ritmo o período do oscilador silábico (cujo correspondente na fala é o intervalo entre *onsets* vocálicos consecutivos) é a unidade mínima de produção da temporalidade.

Os dois experimentos descritos no presente capítulo tem por objetivo produzir evidência comportamental para a hipótese segundo a qual os falantes são sensíveis às mudanças sucessivas na duração dos intervalos definidos pelos *onsets* vocálicos ou unidades V-V.

A abordagem unificadora proposta na seção 1.4 não é exatamente *sui generis*. As citações que seguem mostram como a idéia de um entrelaçamento entre a atividade de produzir e perceber a fala de uma forma geral e o ritmo em particular tem precedentes entre os linguistas.

> predicting the location of upcoming accent plays an important role in sentence comprehension in that it permits a processing mechanism to direct attention to the 'potentially most important elements' of an utterance. [...] The attention directing mechanism is itself guided by the temporal organization of the sentence (Cutler e Foss 1975).

> The perceiver expects as input an intact (as spoken) utterance and has an internal representation of the production constraints and temporal redundancy in the signal. Such constraints and temporal redundancy allow future pattern elements to be predictable in real time (Martin 1986)

> the speaker uses prosody to direct and control aspects of the listener's perception. [...] The speaker can be shown to be monitoring prosody and adjusting it with the listener's comprehension in mind (Cutler 1987).

> Using suprasegmental cues the listener can extract the linguistic structure of the message. Pauses, for example, divide the continuous speech signal into (linguistically) coherent word groups or chunks of acoustic-phonetic information [...]. Hence, adequate pauses increase the intelligibility of the speech signal. Likewise, *accentuation* [grifo do autor] guides listener's attention to the words which are considered important by the speaker, and which are acoustically most reliable (Quené e Kager 1993).

> listeners focus their attention to time points in the speech signal when salient stressed syllables are likely to occur, with this expectancy derived from the rhythmic properties of the speech signal heard so far. (Quené e Port 2005)

O argumento subjacente a estas proposições é que se a produção da estrutura temporal é regulada por algum princípio organizador, então a percepção também deve ser. Se os falantes produzem consistentemente a fala de uma determinada maneira, então é possível supor que os ouvintes se beneficiem desta regularidade de alguma forma. O corolário, interessante do ponto de vista experimental, que se pode deduzir destes argumentos é que se houver algum tipo de regularidade na produção de proeminências na fala, então a percepção do material linguístico que ocorre nestes pontos deve ser facilitada de alguma maneira.

Aqueles que decidiram perseguir a idéia de um acoplamento entre a produção e a percepção da temporalidade na fala investiram em buscar evidência para a facilitação de certos processos de percepção, como a detecção de um determinado fone (Connine e Titone 1996) ou de um tom puro numa frase quando o alvo a ser detectado encontra-se em pontos ritmicamente importantes. Quené e Port (2005) resenham, entre outros, dois trabalhos considerados por eles importantes marcos na exploração da hipótese de facilitação rítmica por conseguirem prover

alguma evidência positiva em seu favor, ao mesmo tempo em que apontam alguns dos obstáculos metodológicos que impediram uma corroboração mais substancial das hipóteses em jogo. O primeiro trabalho resenhado é Meltzer et. al. (1976), no qual relata-se um experimento em que os participantes ouviam frases como a do exemplo 9 e eram instruídos a responder, pressionando um botão, quando ouvissem o fone [b].

(9) The salesman was tired of carrying a heavy briefcase around all day.

O alvo aparecia em três condições: sem manipulação, adiantado e atrasado. A condição com o alvo em posição adiantada foi produzida por meio da retirada de um intervalo de 100 ms de sinal acústico que precedia em 50 ms o início acústico do fone-alvo. A condição com o alvo atrasado foi produzida pela inserção de um fragmento de ruído branco com duração de 200 ms no instante que precedia em 50 ms o início acústico do fone-alvo. Não são mencionadas no artigo original as razões para a escolha dos valores utilizados para manipular a estrutura temporal das frases empregadas no experimento. Aparentemente não houve um estudo piloto que indicasse serem os desvios introduzidos pelos experimentadores na organização temporal dos enunciados suficientes para perturbar o processamento normal de sua estrutura temporal original ou mesmo se eram excessivos a ponto de criar um artefato acústico não processável do ponto de vista linguístico. Esse detalhe é suficiente para alertar sobre a necesidade de interação entre a pesquisa sobre a produção e a percepção da estruturação temporal da fala.

Os valores médios de tempo de reação (TR) mostram que os alvos em posição não manipulada foram detectados mais rapidamente (700 ms) do que os alvos em condição adiantada e atrasada (781 e 765 ms, respectivamente). Estes resultados positivos devem ser considerados com cautela, pois o pouco controle que os experimentadores exerceram sobre aspectos importantes da organização prosódica dos enunciados utilizados no experimento não permitiu que a condição *ceteris paribus* fosse alcançada, de modo que os resultados não permitem afirmar com a clareza desejada em que medida apenas as possíveis expectativas temporais induzidas pela organização duracional do enunciado são responsáveis pelas diferenças entre as médias de TR verificadas no experimento. Outros fatores, como como a divisão do material linguístico em constituintes ou a taxa de elocução, entre outros, podem ter contribuído de forma não planejada para o resultado final.

O segundo experimento importante resenhado por Quené e Port (2005) é o realizado por Shields et. al. (1974), no qual foram comparados os TR para alvos localizados em dois tipos de condição. Na primeira o alvo está localizado, segundo os autores do experimento, numa posição em que o ouvinte esperaria encontrar um elemento forte, dada a estrutura da frase. Na outra, o alvo está localizado numa posição não favorecida do ponto de vista da sua previsibilidade a partir do contexto temporal. No exemplo 10, a estrutura da frase, por hipótese, induz o ouvinte a esperar uma palavra com padrão de acento trocaico no lugar em que o logatoma "benkik" aparece. Portanto, se o alvo fosse o fone [b] (caso da primeira condição), então os TR deveriam ser menores do que a situação em que o alvo fosse o fone [k] (segunda condição).

(10) You will have to curtail any sightseeing plans, as the plane to benkik/benkik leaves at noon.

Os resultados indicam que as médias de TR das duas condições diferem significativamente entre si, quer o logatoma apareça mais no início ou mais no fim da frase (594–675 ms e 537-

640 ms, respectivamente). Haveria, portanto, uma facilitação na tarefa de detecção do alvo induzida pelo contexto. Para mostrar que a diferença nos TR pode ser atribuída exclusivamente à criação de expectativas quanto à ocorrência de uma sílaba acentuada (forte do ponto de vista métrico/rítmico) e não a uma facilitação atribuível à maior ou menor saliência acústica das sílabas portadoras do alvo em função da realização do acento lexical, os logatomas foram inseridos em listas de palavras e a diferenças nos TR entre os alvos previsíveis e imprevisíveis desapareceu.

Apesar dos resultados positivos em favor da hipótese da expectativa temporal, como no caso do experimento anterior, há críticas metodológicas a serem feitas. A intervenção dos experimentadores neste trabalho atuou no sentido de explorar a formação de expectativa métrica mais do que rítmica. As frases utilizadas no experimento controlaram mais **o quê** (regularidade métrica) os ouvintes deveriam esperar (sílaba acentuada ou átona) do que **quando** (regularidade rítmica) os ouvintes deveriam esperar a ocorrência de algum evento. Fica patente, portanto, a importância de investigar a produção da temporalidade na fala como subsídio para a pesquisa de sua contraparte perceptual. Sem saber com maior precisão como os falantes modulam a duração dos fones ao longo de um enunciado para produzir proeminências é muito difícil estabelecer se os ouvintes tiram partido das possíveis regularidades que regem esta modulação. O efeito perturbador introduzido pelas manipulações temporais, que se revela nos experimentos comentados nos parágrafos anteriores pelo aumento nos tempos de detecção dos fones-alvo deslocados temporalmente pode ser interpretado nos quadros da proposta de abordagem integrada da produção e da percepção do ritmo descrita na seção 1.4. Assumindo-se a proposta, a estranheza causada pela antecipação ou o adiamento do fone-alvo, reportada por alguns dos participantes dos experimentos como uma mudança brusca na taxa de elocução, seria consequência de uma perturbação local inesperada no padrão de alongamentos sucessivos das unidades V-V que se espera ouvir ao longo de um grupo acentual. No caso do PB o padrão presente na produção e, por hipótese da abordagem integradora, esperado pelos ouvintes, são alongamentos exponenciais na duração das unidades V-V entre dois acentos frasais (cf. equação 1.1 da função de sincronismo na página 1.1).

5.1 Experimento de percepção I

A hipótese principal a ser posta em teste pelo experimento proposto aqui pode ser sumarizada na seguinte proposição: se existe acoplamento entre produção e percepção da temporalidade na fala, então mudanças no plano da produção devem provocar mudanças no plano da percepção, no sentido de criar expectativas temporais. Há em relação a esta hipótese o pressuposto segundo o qual as mudanças no plano da produção que importam considerar são as mudanças na duração das unidades V-V provocadas pelo mecanismo de acoplamento entre os osciladores acentual e silábico. Se a hipótese está correta, então o ouvinte deve ser sensível ao aumento gradiente da duração das unidades V-V, isto é, dos intervalos entre *onsets* vocálicos consecutivos, ao longo do grupo acentual e, portanto, deve haver um aumento gradiente do nível de atenção ao longo do grupo acentual, que se refletiria numa facilitação da percepção *onsets* vocálicos em posição prosódica forte.

A tarefa experimental proposta aqui para testar essa hipótese é medir o nível de atenção

dirigida pelos ouvintes às sucessivas unidades V-V ao longo de um grupo acentual por meio do monitoramento de tons puros associados ao início de cada unidade V-V ao longo de um grupo acentual. A previsão empírica derivada da hipótese principal é que o TR, isto é, o tempo gasto para perceber a ocorrência do tom, deve ser menor quanto mais próxima da culminância do acento frasal (ponto de máxima influência do oscilador acentual sobre o silábico) estiver a unidade V-V a que o tom está associado.

5.1.1 Estímulos experimentais

As frases que serviram de base para a elaboração dos estímulos do experimento de percepção I foram retiradas do *corpus* do estudo de produção I. Foram escolhidas oito palavras, quatro com duas pré-tônicas (*aratacas*, *patarata*, *horrorosos* e *gonococo*) e quatro com três pré-tônicas (*arapabacas*, *jaratacaca*, *hodofobias* e *locomotiva*), quatro delas começadas por consoante e quatro por vogal.

Selecionaram-se duas versões das frases-veículo para cada palavra, "A *palavra-chave* parece menor hoje" e "A *palavra-chave* bicolor parece menor", equivalendo aos valores de variável $d_\sigma = 0$ e $d_\sigma = 4$. No caso das frases com $d_\sigma = 0$ foram escolhidas as repetições com menos proeminência inicial e no caso das frases com $d_\sigma = 4$ foram escolhidas as repetições em que a proeminência inicial era mais pronunciada.

Os estímulos foram produzidos por meio da inserção de tons puros com frequência de 2,5 kHz em diferentes pontos do grupo acentual. Abaixo segue um esquema simplificado dos locais onde os tons foram inseridos na frase "*A patarata* parece menor hoje". O símbolo $*$ indica a posição do tom puro no grupo acentual e a unidade V-V portadora do acento frasal está destacada em negrito.

- inicial: ap$*$ at ar **at** ap
- pré-tônica 1: ap at$*$ ar **at** ap
- pré-tônica 2: ap at ar$*$ **at** ap
- frasal: ap at ar **at**$*$ ap

Os rótulos dados às posições na lista acima são, numa primeira aproximação, descritivos. Dado o desconhecimento a respeito do comportamento dos sujeitos na tarefa proposta pelo experimento e de seu *span* perceptual não é possível descartar de antemão, por exemplo, a hipótese de que os TR ao clique que está na posição que foi denominada acima de "pré-tônica 2" sofrerão uma influencia forte da unidade V-V seguinte, o V-V /at/ frasalmente acentuado.

No caso da frase "*A patarata bicolor* parece menor hoje" todas as posições são iguais, a menos da posição frasal, como indicado abaixo.

- frasal: ap at ar **at** ab ic ol **orp**$*$

As frases contendo palavras com três pré-tônicas seguiram o mesmo esquema, com a adição da posição pré-tônica 3, conforme se vê abaixo. O exemplo é para as frases "*A locomotiva* parece menor hoje" ($d_\sigma = 0$) e "*A locomotiva bicolor* ($d_\sigma = 4$) parece menor hoje". O símbolo

$*$ indica a posição do tom puro no grupo acentual e a unidade V-V portadora do acento frasal está destacada em negrito.

- inicial: al$*$ oc om ot **iv** ap
- pré-tônica 1: al oc$*$ om ot **iv** ap
- pré-tônica 2: al oc om$*$ ot **iv** ap
- pré-tônica 3: al oc om ot$*$ **iv** ap
- frasal (d_σ= 0): al oc om ot **iv**$*$ ap
- frasal (d_σ= 4): al oc om ot **iv** ab ic ol **orp**$*$

Os tons foram inseridos nos arquivos de som originais por meio de um *script* do programa Praat, sempre na porção medial da consoante que forma cada unidade V-V. Foram gerados no total 72 arquivos de som contendo um tom cada.

Outras 68 frases foram criadas e gravadas para serem usadas como distratores (*fillers*), cada uma contendo um tom associado a uma unidade V-V escolhida aleatoriamente pelo *script* de Praat citado no parágrafo anterior (os tons foram inseridos sempre na porção medial da consoante que forma cada unidade V-V). Cada frase-veículo foi acrescida de uma frase inicial para evitar que a ocorrência dos tons em posição inicial nos itens experimentais causasse algum tipo de surpresa que interferisse drasticamente sobre o tempo de resposta. Em metade dos itens distratores o tom foi inserido na frase inicial e na outra metade na segunda frase. Nos itens experimentais, obviamente, todos os tons estavam na segunda frase. Essa medida foi tomada como tentativa de evitar criar nos sujeitos algum tipo de associação entre a resposta à pergunta de conteúdo e a ocorrência dos tons na primeira ou na segunda frase.

5.1.2 Variáveis

As variáveis independentes investigadas no experimento são indicadas abaixo. Os nomes entre parênteses serão usados para fazer referência às variáveis e grupos na análise estatística dos resultados, relatados na seção 5.1.5.

1. posição no grupo acentual (fator POSIÇÃO):
   - quatro níveis no caso das palavras com duas pré-tônicas (grupo I):
     inicial;
     pré-tônica 1;
     pré-tônica 2;
     frasal.
   - cinco níveis no caso de palavras com três pré-tônicas (grupo II):
     inicial;
     pré-tônica 1;

pré-tônica 2;
pré-tônica 3;
frasal.

2. variável d_σ(fator ACENTO):

   d_σ= 0;
   d_σ= 4.

3. ataque da primeira sílaba da palavra-chave (fator INÍCIO):

   V, como em "As aratacas [...]";
   C, como em "A patarata [...]".

Como variável dependente definiu-se o tempo de reação (TR).

5.1.3 Procedimentos

As 140 frases foram apresentadas em 4 blocos com 35 frases. Entre os blocos os participantes foram instruídos a fazer uma pausa de descanso. O intervalo entre o final de um estímulo experimental e o início do próxima foi de 2 segundos.

Para garantir que os sujeitos processasem integralmente cada frase e não se concentrassem somente na espera do tom, incluiu-se uma pergunta sobre o conteúdo da frase que exigia atenção ao conteúdo semântico do que estava sendo escutado. A pergunta "A frase que você ouviu fala sobre meios de transporte?", que exigia uma resposta do tipo sim ou não, foi apresentada visualmente. As 68 frases distratoras mais 10 das frases-veículo (as que continham a palavra *locomotiva*) empregaram palavras deste campo semântico.

Os 140 estímulos foram apresentados via fones de ouvido fechados da marca *Philips* por meio do programa DMDX (Forster e Forster 2003). O programa registrou o tempo de resposta ao tom, além do acerto ou erro na resposta à pergunta sobre o conteúdo de cada frase. As respostas foram coletadas por meio de um *joystick* da marca *Microsoft*, modelo *Sidewinder*. Dezoito sujeitos participaram voluntariamente do experiemento, que aconteceu em uma sala sileciosa. Todos assinaram um termo de consentimento livre e esclarecido e em sua maioria eram alunos do IEL no momento em que fizeram o teste. Antes de o teste ser iniciado o experimentador lia em voz alta as instruções para o sujeito, o que garantiu que todos receberam as mesmas orientações.

Além de lidas pelo experimentador, as instruções apareciam em seguida por escrito na tela do computador. Antes do início do experimento foi apresentado aos sujeitos um grupo com 8 itens como forma de treinamento, sendo facultada a repetição até que eles estivessem à vontade com a tarefa. Somente um sujeito solicitou a repetição.

5.1.4 Análise estatística

A análise estatística do experimento consistiu na aplicação de testes de ANOVA de dois fatores, aplicadas separadamente aos grupos de palavras com duas e três pré-tônicas. Quando

houve necessidade de fazer múltiplas comparações o teste *post hoc* Scheffé foi utilizado. O nível de significânica (α) em todos os casos foi fixado em 5%.

Dados de tempo de reação apresentam, em geral, distribuição altamente assimétrica (*skewed*) com cauda à direita e, portanto, grande chance de não seguirem a distribuição normal. Essa característica desencoraja o uso de técnicas paramétricas de inferência estatística, fato que obriga a pensar no uso de algum tipo de transformação que aproxime a distribuição dos TR à normal.

A solução adotada aqui consistiu na utilização de dois tipos de transformação que se mostram interessantes como saída para o problema da não-normalidade em dados de duração. Assim como no caso da análise de duração do experimento de produção, aqui empregou-se a logaritmização dos valores brutos e posteriormente a aplicação de z-score, segundo a equação 5.1.

$$z_{ij} = \frac{y_{ij} - \bar{Y}_j}{s_j}, \tag{5.1}$$

em que y_{ij} é a i-ésima amostra do j-ésimo sujeito, $\bar{Y}_j$ é o valor médio das amostras do j-ésimo sujeito e s_j é o valor do desvio-padrão das amostras do j-ésimo sujeito.

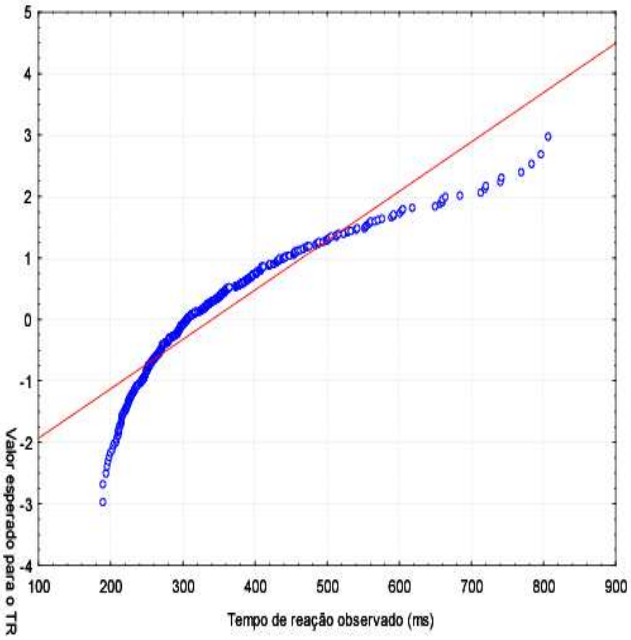
Isso resulta numa série de listas L_j contendo os valores normalizados das respostas de cada sujeito. A amostra contida em cada L_j tem média nula e desvio-padrão unitário (i. e., igual a 1). Da lista de amostras normalizadas foram ignorados os valores menores do que -2 ou maiores do que 2 z-scores, por serem prováveis candidatos a *outlier*. Este procedimento justifica-se pelo fato de 99% dos valores de uma amostra normalmente distribuída com média nula e desvio-padrão unitário situarem-se no intervalo $\pm 1,96$ z-scores. Assim, a probabilidade de algum valor legítimo estar fora do intervalo mencionado é muito baixa, podendo-se com justiça suspeitar que os dados fora dele sejam *outliers*.

Outra razão para a aplicação da transformação z-score é que o valor real dos tempos de reação não é, a princípio, o que mais interessa na análise e sim as diferenças de TR entre as diversas condições experimentais. Como o z-score mede o grau de afastamento relativo de um dado em relação à média da sua amostra, a transformação por z-score mostra-se adequada para o caso em questão.

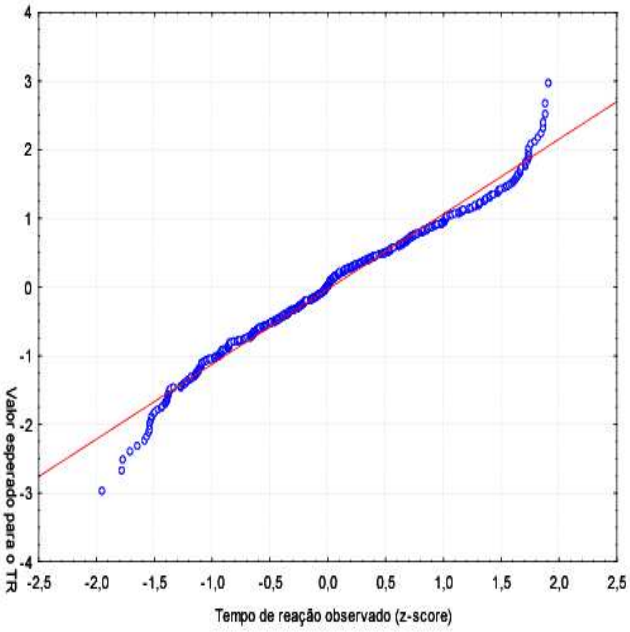
A figura 5.1(a) mostra o gráfico de probabilidade observada versus esperada (normal) para os dados brutos (em ms) de TR do grupo de palavras com duas pré-tônicas. Se a distribuição fosse perfeitamente normal, todos os pontos deveriam estar sobre a linha contínua. A figura 5.1(b) mostra o gráfico de probabilidade normal para os dados normalizados do mesmo grupo. Note-se como os dados normalizados seguem mais de perto os valores normais esperados para a amostra.

5.1.5 Resultados

A ANOVA de 3 fatores (POSIÇÃO, ACENTO e INÍCIO) realizada nos dados de TR do grupo I indicou que apenas o efeito do fator POSIÇÃO é significativo ($F(3, 435) = 10,54\ p < 10^{-5}$). Nenhuma das interações alcançou significância. A comparação entre as posições individuais revela que as posições pré-tônica 1 e 2 não diferem entre si, a inicial difere marginalmente (Scheffé, $p = 0,063$) da pré-tônica 1, sendo significativas as demais comparações com pelo



(a) Gráfico de probabilidade normal dos dados brutos de TR



(b) Gráfico de probabilidade normal dos dados normalizados de TR

Figura 5.1

menos $p < 0,03$. As figuras 5.2(a) e 5.2(b) mostram os valores médios de TR para cada posição nos grupos I e II.

A ANOVA realizada nos dados de TR do grupo II indica, como no caso anterior, que apenas o efeito do fator POSIÇÃO alcança significância ($F(4,563) = 6,9\ p < 10^{-4}$). De novo, nenhuma interação alcançou significância. As comparações individuais mostram que as posições pré-tônica 3 e frasal são significativamente diferentes das posições inicial e pré-tônica 1 ($p = 0,04$ pelo menos) sendo marginal a diferença em relação à posição pré-tônica 2 ($p = 0,064$). Todas as demais comparações falham em apontar significância.

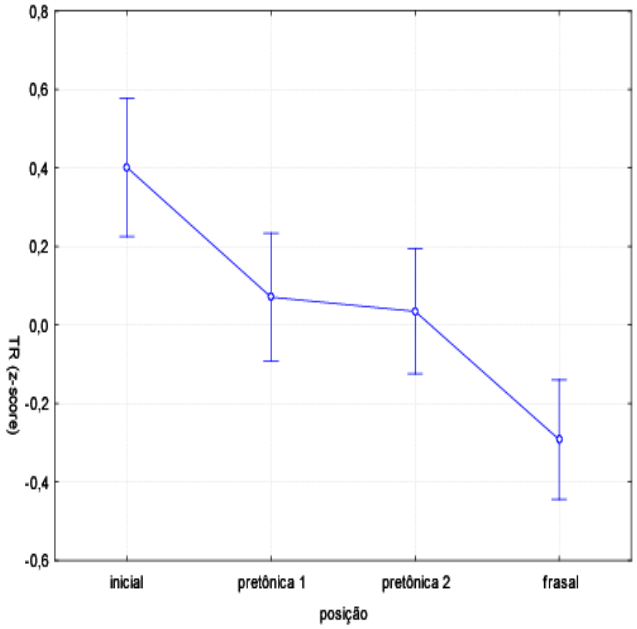
Para por à prova a sensibilidade do desenho experimental às diferenças de duração presentes nas frases, fez-se uma correlação entre a duração das unidades V-V e as médias de TR correspondentes. No caso do grupo I, a correlação ($r = -0,63$) é significativa ($p < 10^{-3}$) e a variabilidade dos dados de duração explicam em torno de 40% da variância dos dados de TR. O fato de a correlação ter sinal negativo indica que quanto maior a duração, menores os TR. Este resultado é um indício de uma ligação entre a variabilidade no parâmetro acústico duração e modificações nos níveis de atenção. O que se pode entrever, no entanto, é que a variação no parâmetro duração deve ter uma magnitude relativamente alta para que seja possível notar mudanças importantes na atenção dispensada a cada unidade rítmica na cadeia da fala.

5.1.6 Discussão

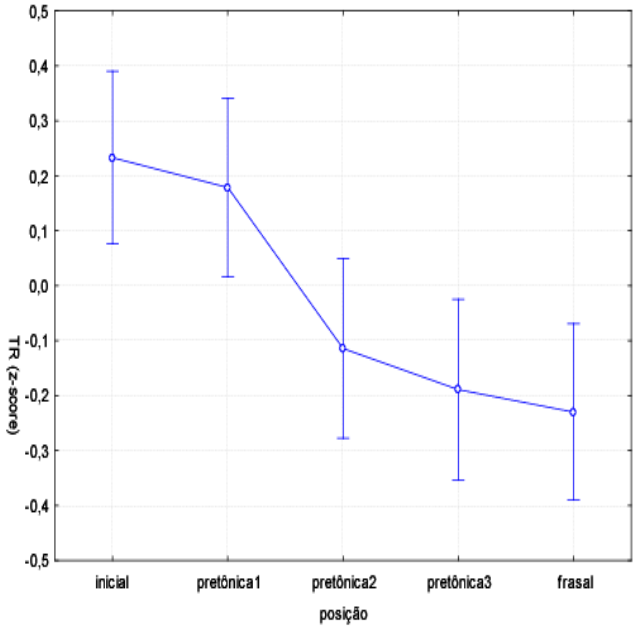
A novidade e o interesse do experimento I é mostrar evidência de uma relação entre a produção e a percepção da estruturação temporal da fala. Tomando como ponto de partida o fenômeno da acentuação secundária, tentou-se angariar evidência para a estreita relação de entrelaçamento entre as duas atividades. Ainda que preliminares, os resultados parecem sugerir que a comunicação falada, naquilo que importa para a prosódia, implica numa atividade de acoplamento entre, de um lado, o falante que produz proeminências e, do outro, o ouvinte que deve seguir as pistas deixadas pelo falante no sinal acústico sobre onde este deve esperar que ocorram os trechos mais importantes da enunciação daquele. Dois resultados parecem apontar nesta direção: (1) a latência de detecção do tom diminui à medida que a proeminência acentual vai se formando, i.e., à medida que a duração das unidades V-V vai aumentando quando se aproxima a culminância do acento frasal e (2) a duração sozinha dá conta de explicar em torno de 40% da variância dos dados de tempo de reação, o que indica a importância da duração para o fenômeno do ritmo da fala. O processo de acentuação, assim, parece ir se formando ao mesmo tempo na enunciação do falante e na percepção do ouvinte.

5.2 Experimento de percepção II

Os resultados do experimento I parecem mostrar que no interior de um grupo acentual o contorno da duração parece guiar a percepção do ouvinte. Entretanto, uma resposta vital para a hipótese de uma ligação fundamental entre a produção e a percepção da temporalidade da fala não pode ser respondida pelos resultados. Qual é comportamento da percepção dos sujeitos na presença de uma fronteira entre dois grupos acentuais ou, colocando em termos do modelo do ritmo, na presença da batida do oscilador acentual? Na produção, os grupos acentuais



(a) Médias de TR normalizado (*z-score*) para cada posição no grupo I.



(b) Médias de TR normalizado (*z-score*) para cada posição no grupo II.

Figura 5.2

constituem unidades coesas de organização do ritmo. Essa coesão é gerada pela combinação do mecanismo de indução do oscilador silábico e pela ação do mecanismo de *reset* que, após a batida do oscilador acentual, faz o oscilador silábico voltar à sua frequência natural. A ação de *reset* tem, assim, o efeito de criar pontos de relativa descontinuidade na estrutura temporal da fala. Daí possivelmente deriva a sensação de uma "quebra" associada às fronteiras. O experimento I, em função do seu desenho, apenas mostrou uma sensibilidade dos sujeitos ao mecanismo de indução ao mostrar que sua percepção é de alguma forma capturada pelo padrão de mudanças subjacente ao contorno duracional presente no sinal acústico escutado por eles. Entretanto, a hipótese de um entrelaçamento entre produção e percepção prevê que os sujeitos tem que ser também sensíveis ao mecanismo de *reset*, isto é, eles precisam ser sensíveis aos momentos em que cessa momentaneamente a ação indutora do oscilador acentual e o oscilador silábico do falante, responsável pela sequência durações no sinal produzido, retorna à sua frequência natural de oscilação.

O experimento II, com o objetivo de tentar responder às limitações do experimento I, introduz uma variação crucial. Essa modificação consiste em fazer os sujeitos detectarem cliques ao longo de dois grupos acentuais consecutivos e não apenas de um. Espera-se reproduzir com o experimento II o que Martin (1986) chamou de *scallop effect*, que será traduzido aqui por efeito lombada. Sucesso em reproduzir este efeito constituiria evidência favorável à hipótese de acoplamente entre produção e percepção da temporalidade da fala. A seção 5.2.1 descreve em mais detalhes o efeito lombada.

5.2.1 O efeito lombada

Em Martin (1986), o autor relata um experimento que é bastante semelhante aos relatados neste capítulo na estrutura geral e nos objetivos. Uma das diferenças é que a tarefa experimental apresentada aos sujeitos é a detecção de fone e não de cliques como é o caso aqui. Outra diferença é que os estímulos são frases formadas por logatomas monossilábicos ao invés de palavras. Um dos estímulos é apresentado no exemplo 11.

(11) LAS_1 a PAS_2 a LAS_3**,** a GAS_4 a LAS_5 a BAS_6.

As frases são compostas por onze monossílabos, sendo seis deles tônicos (representados em maiúsculas em 11 e cinco átonos, possivelmente reduzidos a [ə]. Uma característica importante dos estímulos é que eles foram produzidos de modo a haver uma fronteira prosódica entre o terceiro e o quarto monossílabos tônicos, na posição indicada pela vírgula no exemplo 11[1].

Entre os monossílabos tônicos, há alvos (PAS, BAS, KAS e GAS) e um distrator (LAS). Cada frase tem, portanto, seis posições-alvo, três antes da fronteira e três depois. Metade das frases do experimento tem possíveis alvos nas posições 1, 3 e 5 e outra nas posições 2, 4 e 6. Os participantes ouviam todo o conjunto de frases, mas foram divididos em quatro grupos, cada qual encarregado de identificar um dos possíveis fones-alvo [p, b, k, g], de modo que a cada frase o sujeito deveria fazer apenas uma identificação. A tarefa que eles realizaram foi pressionar um botão assim que identificassem a presença de sua consoante-alvo. Não há informações sobre o

[1] O texto não diz qual foi o critério (se perceptivo ou baseado em alguma análise acústica) adotado para garantir que havia de fato uma fronteira nas frases utilizadas no experimento.

número de participantes ou sobre detalhes da análise estatística dos dados. A figura 5.3 mostra o tempo de reação médio em cada uma das seis posições alvo.

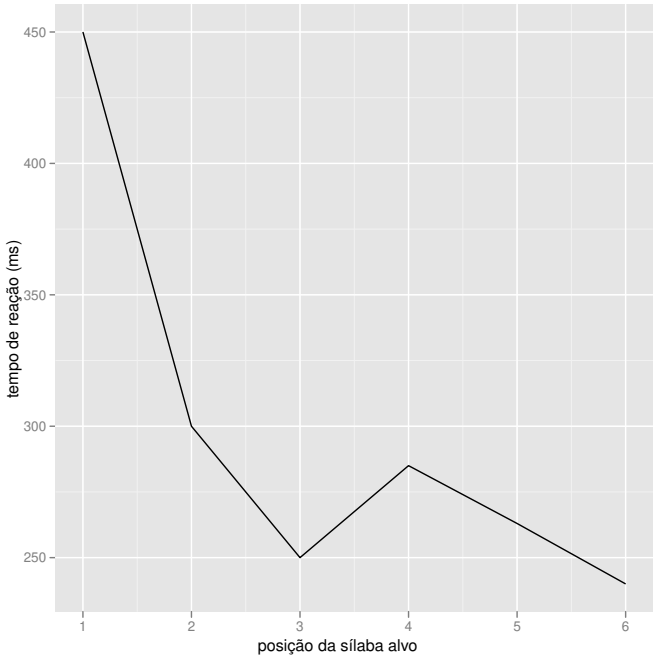


Figura 5.3: Tempo de reação médio para a identificação do fone-alvo em cada posição das frases de teste no experimento de Martin (1986). A fronteira acústica está localizada entre as posições 3 e 4. O gráfico é uma adaptação da figura original.

Pode-se observar que, assim como no experimento I, os tempos de reação no experimento de Martin tendem a ser menores conforme o alvo esteja mais à direita na frase. O elemento crucial na figura, entretanto, é o aumento relativo no tempo de reação da posição 4 em relação à posição 3, que coincide com a fronteira prosódica existente no sinal acústico. Nas posições subsequentes os tempos de reação voltam a cair, como se a fronteira fosse uma lombada que obrigasse os sujeitos a frear antes de passar por ela. A interpretação do autor para os resultados é que a fronteira representa um lugar de descontinuidade no nível da produção e essa descontinuidade é espelhada na percepção. Esta interpretação coaduna-se com a abordagem integradora da produção e da percepção delineada na seção 1.4. Nos termos da proposta delineada ali, a sensação de descontinuidade associada à ocorrência de um acento frasal no nível da percepção pode ser associada à recuperação que o ouvinte faz dos momentos em que ocorre a ação, no nível da produção, do mecanismo de *resetting*, que interrompe o processo de indução de acoplamento de período entre os osciladores silábico a acentual criando, um ponto de descontinuidade por meio do contraste entre a duração longa do último período do oscilador silábico ainda em acoplamento e a duração relativamente mais curta do período do oscilador silábico em repouso após o *resetting*.

5.2.2 Estímulos experimentais

Do *corpus* do experimento de produção I foram selecionadas 8 frases-veículo. As frases escolhidas foram do tipo exemplificado pelo exemplo 1e na seção 3.1.2, reproduzida abaixo.

(12) |A *palavra-chave*| bicolor| parece menor.

Foram escolhidas frases contendo palavras-chave com duas e três sílabas pré-tônicas, definindo dois grupos:

Pr2: patarata, gonococo, arataca e horrorosos (8 posições V-V)

Pr3: jaratacaca, locomotiva, arapabacas e hodofobias (9 posições V-V)

Das 10 repetições de cada frase contendo as palavras escolhidas que estavam disponíveis escolheu-se uma versão em que havia a presença de uma fronteira (segundo os critérios para delimitação de grupos acentuais descritos no apêndice A) entre a palavra-chave e o adjetivo *bicolor*, da forma como ilustra o exemplo 12. A figura 5.4 mostra algumas das características acústicas da repetição da frase "A patarata bicolor parece menor" escolhida para o experimento II. Observe-se o pico de duração na unidade V-V 4, seguida do *reset* de duração na unidade seguinte, definindo a ocorrência de uma fronteira. O contorno de duração normalizada mostrado na figura foi obtido por meio do *script* `SGDetector` (cujo código-fonte pode ser visto em Barbosa 2006, pp. 459–468), que é uma implementação do algoritmo descrito no apêndice A. O contorno de f_0 mostrado na figura foi obtido pela função *Periodicity* do Praat ao qual aplicou-se posteriormente a função de suavização do Praat (5 Hz de largura de banda).

Foram feitas oito cópias do arquivo de som de cada estímulo do grupo Pr2 e nove dos estímulos do grupo Pr3. Um clique foi inserido em cada arquivo em cada posição V-V ao longo dos dois grupos acentuais de interesse. Dentro de cada unidade o clique foi preferencialmente inserido após a vogal, em geral no ponto médio de duração dos segmentos consonantais pós-vocálicos.

5.2.3 Procedimento experimental

O experimento consitiu na apresentação de 67 itens experimentais [(4 frases × 8 posições V-V) + (4 frases × 9 posições V-V)][2] mais 73 frases distratoras. A apresentação dos 140 estímulos, que se sucediam a intervalos de 800 ms, foi dividida em 4 blocos, entre os quais os sujeitos puderam fazer uma pausa. O intervalo entre a apresentação dos estímulos foi de 800 ms. Os estímulos foram apresentados por meio do programa DMDX (Forster e Forster 2003) via fones de ouvido. Os tempos de reação foram coletados através de um mouse da marca Logitech modelo G5, com uma taxa de atualização de 500 Hz. Quarenta e dois sujeitos falantes nativos de português brasileiro sem nenhum problema auditivo reportado e com idades em 20 e 35 anos participaram do experimento, completando o experimento em torno de 20 minutos.

Para tentar garantir que os sujeitos não ignorassem a frase e apenas monitorassem o ocorrência do clique, adotou-se o mesmo procedimento usado no experimento I e descrito na seção 5.1.3 de fazer uma pergunta de conteúdo a respeito de cada estímulo após sua apresentação.

[2]Em função do encontro vocálico na palavra *hodofobia* perde-se uma posição V-V na contagem total dos itens experimentais.

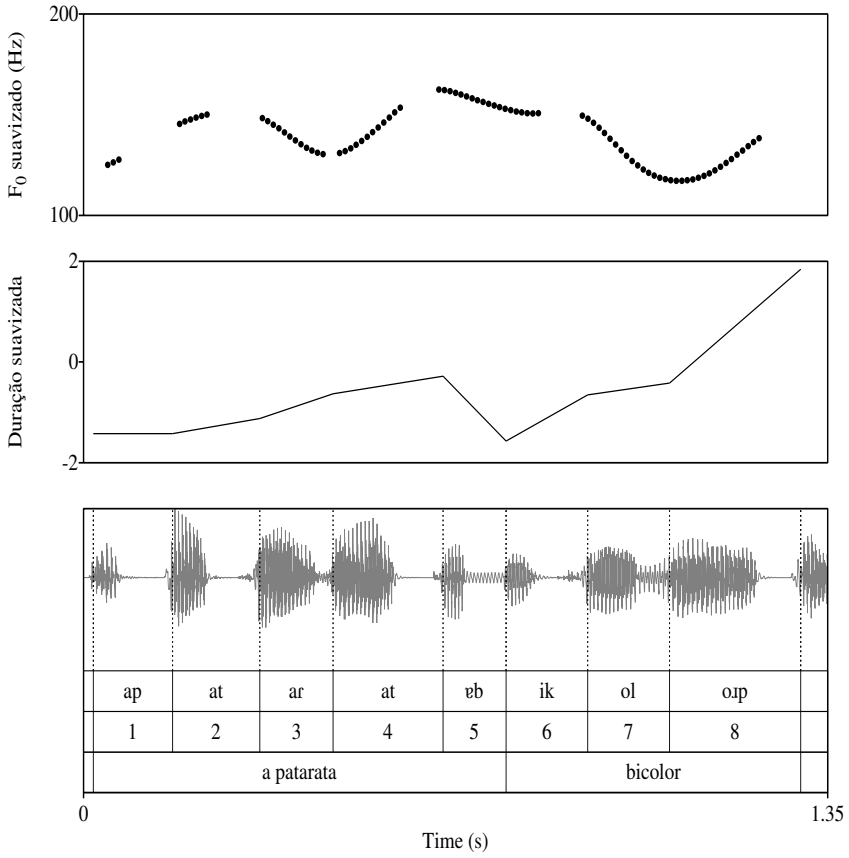


Figura 5.4: Caracterização acústica de um dos estímulos do experimento II. No painel superior mostra-se a curva suavizada da f_0. No painel central o contorno da duração normalizada e no inferior a forma de onda do trecho relevante da frase-veículo. As oito unidades V-V que formam os dois grupos acentuais estão delimitadas pelas linhas verticais pontilhadas. A fronteira entre os dois grupos é definida pelo pico de duração na unidade 4.

5.2.4 Análise estatística

A variável independente no experimento são as posições alvo ao longo dos dois grupos acentuais. Oito no caso das frases do grupo Pr2 e nove no caso das frases do grupo Pr3. A variável dependente é o tempo de reação aos cliques.

Os dados de tempo de reação receberam tratamento semelhante ao descrito na seção 5.1.4, embora a etapa de logaritmização anterior à normalização tenha sido substituída pela transformação de Box–Cox (Box e Cox 1964), segundo a equação 5.2. O valor ótimo de λ para os dados do experimento II, obtido por um procedimento automático, foi estimado em 0,8.

$$y_i^{(\lambda)} = \frac{y_i^{\lambda} - 1}{\lambda} \tag{5.2}$$

5.2.5 Resultados

A figura 5.5 mostra o valor de tempo de reação normalizado médio de cada uma das 8 posições-alvo. A inspeção da figura mostra uma diminuição progressiva nos tempos médios de reação, com um aparente aumento local na posição 6. A fronteira prosódica marcada pela duração encontra-se na transição entre as posições 4 e 5.

A ANOVA aplicada aos dados do grupo Pr2 mostra uma alta significância do fator POSIÇÃO $[F(7, 1170) = 17,830,\ p < 0,001]$. A princípio pode-se esperar que, se o aumento no tempo de reação observado na posição 6 for uma manifestação do efeito lombada, a média dessa posição deve então ser diferente da médias na sua vizinhança imediata. As comparações pareadas (com p corrigido pelo método de Holm) mostram não haver diferença significativa entre as médias das posições posições de 4 a 8, resultado que a princípio não corroboraria a hipótese. No entanto, embora a média da posição 6 não seja diferente de sua vizinhança imediata, ela também não difere significativamente das médias das posições 2 e 3. Estas duas últimas, por sua vez diferem em conjunto de todo o conjunto formado pelas posições 4 a 8 ($p < 0,02$), com exceção da posição 6. Apesar de não inequívoco, o aumento no tempo de reação na posição 6 parece ser motivado pela presença da fronteira. O fato de a "lombada" estar atrasada de uma unidade em relação à fronteira pode ter duas possíveis explicações: (1) o atraso pode ser efeito de *spillover*[3], o que é comum em experimentos psicolinguísticos ou (2) o efeito da presença da fronteira pode acontecer não após a sílaba frasalmente acentuada, mas sim após o final da palavra fonológica, que incluiria também a sílaba pós-tônica.

A figura 5.6 mostra o valor de tempo de reação normalizado médio de cada uma das 9 posições-alvo. Pode-se observar na figura que a queda nos tempos de reação nas quatro

[3] *Spillover* é o termo usado por psicolinguistas para fazer referência a situações em que o efeito observado na variável dependente de um experimento atribuível a uma variável independente em um determinado trecho do estímulo experimental é devido não apenas à carga de processamento exigida pelo trecho corrente, mas também pelo processamento incompleto de um trecho anterior que persevera ou deborda-se no trecho seguinte. Mitchell (1984, p. 76) define *spillover* da seguinte maneira: "In most immediate processing tasks the end of one response measure is immediately followed by the beginning of another, together with a new portion of text. In this situation any uncompleted processing will spill over from one response measure to the next. In others words, certain aspects of processing will be postponed and join a queue or buffer so that they can be dealt with later. [...] Here, the response measure will be influenced not only by the problems in the current display but also by any backlog or processing that may have built up in the buffer."

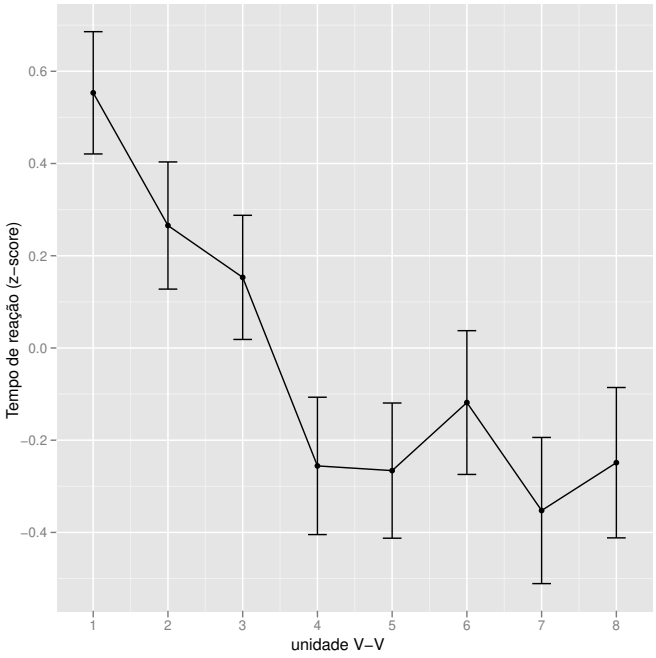


Figura 5.5: Tempo de reação normalizado médio de cada uma das 8 posições-alvo no grupo Pr2. Linhas verticais indicam intervalo de 95% de confiança em torno da média.

primeiras posições não é tão acentuada quanto a que acontece no grupo Pr2. Da quinta posição em diante as médias caem de forma mais visível, muito embora a posição 6, à semelhança do que se observou no grupo Pr2, pareça também sinalizar um pico local.

A ANOVA aplicada aos dados do grupo Pr3 mostra que o fator POSIÇÃO tem influência significativa sobre o tempo de reação [$F(8, 1270) = 11,3,\ p < 0,001$]. Os resultados das comparações pareadas (com p corrigido pelo método de Holm) parecem favorecer a interpretação de que a média da posição 6 é um pico local. Todas as comparações envolvendo as posições 5, 7, 8 e 9 não atingem significância, bem como a comparação entre 5 e 6. Entretanto, as comparações 6–7 e 6–9 resultam significativas ($p < 0,02$). No grupo Pr3 a "lombada" também está alinhada à posição 6, neste caso a unidade V-V que contém a sílaba pós-tônica. Esse resultado torna mais provável que o atraso observado no grupo Pr2 seja causado por *spillover*.

Os resultados da análise das diferenças entre as médias de tempo de reação ao longo dos grupos acentuais parecem reproduzir com algum nível de sucesso o efeito lombada de Martin (1986), trazendo evidência favorável à hipótese de uma interação entre os mecanismos de produção e percepção. Os resultados, no entanto, não deixam claro quais são os parâmetros acústicos presentes no sinal acústico que mais colaboram para a produção do efeito.

Na tentativa de esclarecer essa questão foram feitos testes de regressão linear simples e múltipla entre os valores médios de tempo de cada uma das 67 posições-alvo com vários parâmetros acústicos extraídos dos estímulos do experimento. O teste de regressão simples que alcançou o melhor coeficiente de determinação ($R^2 = 0,4085\ p < 0,001$) foi o que envolveu o parâmetro duração suavizada da unidade V-V$_{i-1}$, isto é, a duração da unidade anterior à posição do clique

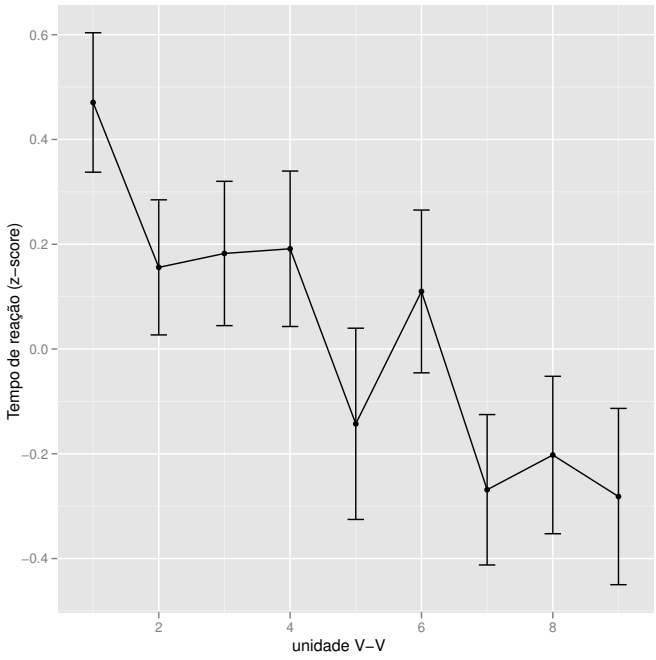
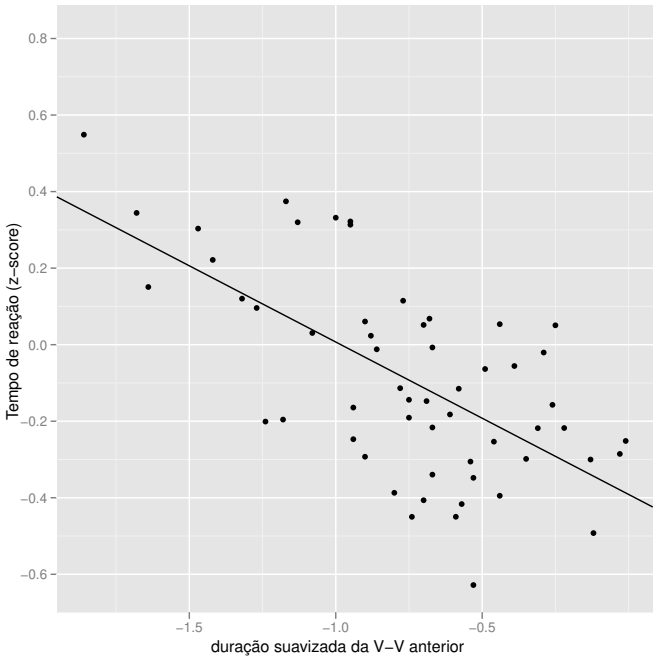
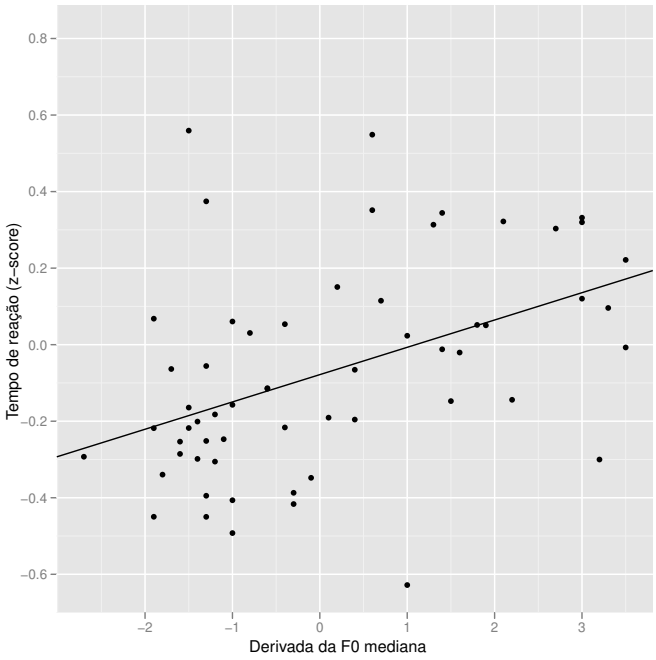


Figura 5.6: Tempo de reação normalizado médio de cada uma das 8 posições-alvo no grupo Pr3. Linhas verticais indicam intervalo de 95% de confiança em torno da média.



(a) Duração



(b) Frequência fundamental

Figura 5.7: Correlação entre parâmetros acústicos e tempo de detecção dos cliques

sendo detectado. Este parâmetro, sozinho, explica em torno de 40% da variância dos dados de tempo de reação. A regressão múltipla que alcançou melhor R^2 envolveu o mesmo parâmetro de duração em conjunto com a derivada discreta do valor mediano da f_0, isto é, a diferença entre o valor mediano de f_0 (medido em semitons re 100 Hz) da posição atual e o valor deste parâmetro na posição imediatamente anterior. O resultado dessa regressão é $R^2 = 0,501$ $p < 0,001$.

A figura 5.7a mostra a dispersão dos dados de tempo de reação em função da duração suavizada da unidade V-V anterior. A linha reta representa a reta da função linear que melhor se ajusta aos dados e mostra que existe uma correlação negativa entre as variáveis, o que significa que o tempo de detecção do clique numa dada posição-alvo tende a ser menor quanto maior tenha sido a duração da unidade imediatamente anterior.

A figura 5.7b mostra a dispersão dos dados de tempo de reação em funçao da derivada discreta da f_0 mediana. A reta de correlação indica uma correlação positiva entre as duas variáveis. Portanto, o tempo de detecção de um clique na posição atual será maior quanto mais positiva seja a diferença entre o valor mediano da posição atual em relação ao mesmo parâmetro na posição imediatamente anterior.

5.3 Discussão

Tomados em conjunto, os resultados dos dois experimentos produziram evidência que pode ser interpretada como favorável à hipótese de uma ligação importante entre a produção e a percepção do ritmo da fala. A reprodução do efeito lombada, ao mostrar que a fronteira entre dois grupos acentuais presente no sinal acústico provoca um efeito no comportamento da atenção dos ouvintes, indica que o grupo acentual é a unidade organizadora da temporalidade da fala tanto do ponto de vista da produção quanto da percepção. Do ponto de vista do modelamento essa evidência empírica é importante, uma vez que sugere que se as duas modalidades de comportamento estão relacionadas então é possível usar o mesmo tipo de modelamento que foi aplicado com sucesso à produção para dar conta da percepção do ritmo.

Um modelo da percepção da temporalidade baseado em osciladores foi proposto por McAuley (1995) e inspirou a formulação do modelo de Barbosa (2006) que tomamos como referência nesta tese. O modelo de McAuley recorre à idéia de acoplamento entre um oscilador perceptual interno e os estímulos ambientais para explicar o comportamento dos sujeitos em tarefas que envolvem a percepção da duração de intervalos de tempo. McAuley não tenta usar seu modelo para dar conta da percepção da estruturação temporal da fala e o trabalho de Barbosa (2006) sugere que o modelo não seria capaz de fazê-lo, pois, assim como proposto, o modelo de McAuley possui apenas um oscilador interno e para modelar a complexidade temporal da fala (pelo menos no caso do português brasileiro) são necessários no mínimo dois osciladores internos que se influenciem mutuamente. Entretanto, Barbosa (2006) sugere que seu próprio modelo pode ser estendido para dar conta da percepção se incorporar a idéia de um estímulo externo (o sinal de fala) que induz um oscilador silábico perceptual no ouvinte. O ouvinte usaria a informação sobre os alongamentos sucessivos de período desse oscilador silábico interno ao longo de um grupo acentual para estimar a força de acoplamento entre os osciladores no sistema de produção do falante. Ao detectar máximos na função de sincronismo na fala do locutor o ouvinte poderia reconstruir no seu sistema perceptivo o lugar aproximado da batida

do oscilador acentual no sistema de produção do falante.

No geral, tanto a análise do tempo de reação médio em cada posição-alvo nos grupo acentuais quanto os resultados dos testes de regressão dos tempos de reação com os parâmetros acústicos mostram que os ouvintes são sensíveis às mudanças nos parâmetros acústicos da fala. É interessante observar que nos testes de regressão os melhores resultados são conseguidos quando são usadas medidas *relacionais*. No caso do parâmetro duração, o ouvinte parece ser sensível à *comparação* entre a duração da posição sendo correntemente monitorada e a duração da posição anterior. O mesmo acontece com a f_0: a variação na f_0 que mais tem impacto nos tempos de reação é a diferença entre os valores medianos de uma posição para outra.

A obtenção de resultados empíricos como os descritos no presente capítulo que mostram ser a percepção que os ouvintes tem da sucessão de *onsets* vocálicos afetada pela maneira como esses intervalos são estruturados ao longo da frase são um passo preliminar para a corroboração da abordagem integradora da produção e percepção do ritmo sugerida na seção 1.4. Três variações dos experimentos de detecção de cliques descritos neste capítulo podem trazer ainda mais confirmação para o uso das unidades V-V como unidade básica da percepção do ritmo:

1. Manipulação da força da fronteira prosódica para verificar a relação desta variável com o tamanho do efeito lombada. A força da fronteira poderia ser manipulada de pelo menos duas formas:

   (a) criação de diferentes versões das sentenças-veículo em que a fronteira sintática separando os grupos acentuais de interesse tenha forças distintas (nesse caso se colocaria o problema da métrica a ser utilizada para medir a força dessa fronteira sintática);

   (b) Manipulação artificial da magnitude da fronteira via técnicas de análise resíntese do contorno duracional da frase-veículo.

2. Replicar o experimento de percepção II com estímulos delexicalisados por meio da aplicação de filtro passa-baixa para estimar quanto o comportamento dos sujeitos na tarefa de detecção de clique é afetado pelo conhecimento léxico-gramatical dos sujeitos.

Como foi dito na seção 1.4 são necessários ainda experimentos que mostrem ser possível que os sujeitos façam, a partir da percepção dos *onsets* vocálicos, estimativas do período do oscilador silábico em repouso e da força de acoplamento entre os osciladores acentual e silábico. O uso de fala delexicalizada, como na variação 2 sugerida acima, pode servir também para estabelecer o impacto no desempenho na detecção dos acentos frasais da falta de uma estimativa, por parte do ouvinte, da força de acoplamente entre o oscilador acentual e a sintaxe.

Conclusão

E eu ao contrário encontro o deleite mais jubiloso em desenredar uma bela e intrincada intriga. E será ainda porque no momento em que, como filósofo, duvido que o mundo tenha uma ordem, consola-me descobrir, se não uma ordem, pelo menos uma série de conexões em pequenas porções dos negócios do mundo.

Umberto Eco, *O nome da rosa*

Dois núcleos temáticos foram investigados ao longo da tese, ambos relacionados com o modelo dinâmico do ritmo (Barbosa 2006) apresentado no capítulo 1. O primeiro deles é a descrição fonético-acústica do acento secundário em português brasileiro com vistas à obtenção de dados que possam estabelecer se o modelo de ritmo, em sua atual formulação, produz contornos de duração adequados quando comparados com os dados empíricos. O segundo tema tratado na tese são as possibilidades que o modelo dinâmico do ritmo e a teoria dos sistemas dinâmicos oferecem para a formulação de uma abordagem que integre a produção e a percepção do ritmo da fala em um mesmo modelo.

Em relação ao primeiro núcleo temático, uma das principais conclusões a que os dados relatados nos capítulos 3 e 4 permitem chegar é que não se verifica nos contornos de duração observados nos experimentos nenhum padrão de alternância binária e que, portanto, a atual formulação do modelo, com apenas dois osciladores, um acentual e um silábico, é adequada para explicar a produção da estruturação temporal da fala.

Além de contribuir para testar a adequação da formulação do modelo do ritmo com dois osciladores e avaliar se a binaridade é o padrão de manifestação privilegiado de acentos secundários, os dados dos dois experimentos de produção propõem uma descrição fonética que pode servir de base para a proposição de outras explicações para para o fenômeno. A análise do parâmetro acústico duração mostra que uma proeminência secundária pode manifestar-se na forma de um alongamento inicial observado principalmente em casos de início absoluto de enunciado que foi chamada de proeminência inicial nos capítulos 3 e 4. Outros parâmetros acústicos além da duração também manifestam um padrão que pode ser descrito como proeminência inicial, como é o caso da f_0 e da abertura vocálica. No caso da f_0, observa-se a presença de picos

de variação alinhados ao início do grupo acentual que poderiam ser percebidos como um tom alto nesta posição. Os dados de abertura vocálica, inferida pelo valor do primeiro formante das vogais, mostram também um abaixamento relativo da primeira vogal do grupo acentual que, em conjunto com a duração e a f_0, poderia colaborar para a percepção de uma proeminência inicial. A variação do parâmetro ênfase espectral, apesar de parcialmente correlacionada com a da duração, seguindo uma trajetória gradiente de culminação, não apresenta valores relativamente mais altos na posição inicial do grupo acentual. As variáveis sintáticas, introduzidas no experimento de produção II, não provocaram efeitos significativos nos parâmetros duração e f_0 no que diz respeito ao acento secundário, o que quer dizer que a posição da palavra na frase não provoca, sozinha, mudanças importantes na forma do contorno de duração nem na distribuição de picos de f_0 ao longo da cadeia de sílabas pré-tônicas. A variável pragmático-discursiva, a exemplo da sintática, não provocou nos contornos de duração normalizada e em f_0 nenhum efeito significativo. A duração bruta da palavra morfológica, no entanto, indica encurtamento na condição de referente já mencionado no discurso. Este resultado, embora não diga respeito ao acento secundário de forma direta, é interessante na medida em que constitui evidência do uso da prosódia como correlato de uma distinção relevante no nível da organização do discurso.

O conjunto de dados produzidos pelos dois experimentos de produção contribui para o melhor entendimento do acento secundário. Algumas das generalizações que os dados sugerem é que a ocorrência de proeminências que não a frasal nos grupos acentuais é muito variável, dependente de sujeito e possivelmente de níveis linguísticos mais elevados como a pragmática e a organização do discurso. A percepção de que entre sílabas que têm em princípio o mesmo grau de tonicidade, como é o caso das pré-tônicas, é possível criar diferenças relativas de proeminência é a intuição básica que subjaz à proposição da existência dos acentos secundários. Embora seja difícil negar que essa intuição faça sentido, os dados apresentados aqui podem ser usados para repensar o tratamento que vem sendo dispensado ao fenômeno a que se deu o rótulo "acento secundário". Um dos problemas com o rótulo é que, como aponta Bertinetto (1976 e 1981) em sua análise apresentada na seção 2.5.2, ele é usado para fazer referência a fenômenos diferentes. Uma inconveniência de chamar de acento secundário as proeminências que podem ser criadas na cadeia de sílabas pré-tônicas é que isso cria a expectativa de que o comportamento dessas proeminências tem o mesmo caráter de outros fenômenos que recebem o mesmo nome. Assim, falar em "acento secundário" sugere que ele tem uma natureza semelhante à do acento primário ou lexical, que ele seja discretizável a ponto de ter um papel tão forte na fonologia da língua quanto os acentos primários ou lexicais, o que não parece ser o caso. No entanto, abolir o rótulo "acento secundário" pode ser uma atitude tão radical quanto negar que proeminências secundárias existam. Uma alternativa intermediária seria favorecer o uso da expressão "acentuação secundária" para fazer referência à potencialidade que as sílabas pré-tônicas oferecem para a criação de relações de proeminência. O uso do substantivo de base verbal *acentuação* para fazer referência ao fenômeno das proeminências secundárias parece apropriado na medida em que sua própria composição morfológica explicita a ligação que o substantivo guarda com a ação expressa pelo verbo *acentuar*. Seria preservada na palavra *acentuação* a noção de que há alguma coisa em curso em contraste com *acento*, que nomeia um resultado, um ente acabado, delimitável.

Pelo menos duas críticas podem ser feitas aos experimentos de produção reportados na tese. A primeira delas diz respeito ao número limitado de sujeitos gravados, dois. Essa limitação

pode ser de fato um problema dado que, como foi dito antes, a ocorrência de proeminências secundárias tem um caráter variável e tem traços idiossincráticos. No entanto, o número elevado de controles adotado nos experimentos tornou difícil aumentar o número de sujeitos sem tornar impraticável o volume de medidas a serem realizadas. Além disso, a amostra mais limitada de sujeitos foi uma medida de cautela. O número relativamente alto de repetições de cada condição, especialmente no experimento I (dez repetições) permitiu ter ideia da variabilidade intrasubjetiva, que se mostrou não ser muito grande para o sujeito daquele experimento. Sem ter ideia dessa variabilidade, ter optado de início por um número menor de repetições para gravar mais sujeitos poderia ter prejudicado as análises estatísticas caso a variabilidade intrasubjetiva fosse alta.

A segunda crítica é o questionamento da possibilidade de generalização dos resultados obtidos aqui, baseados em análises de fala lida, para outros estilos de fala. Não é possível dizer *a priori* se a resposta ao questionamento é afirmativa ou não, já que os dados sobre a ocorrência de proeminências secundárias em outras modalidades são muito escassos.

O segundo núcleo temático da tese trata da questão da construção de uma abordagem integrada da produção e da percepção do ritmo da fala. Os dois experimentos de percepção reportados na tese foram realizados com o objetivo de produzir evidência favorável para a hipótese segundo a qual os intervalos entre *onsets* vocálicos, que são, no nível da produção, a unidade mínima do planejamento do ritmo, desempenham, na percepção, papel semelhante. Evidência favorável a esta hipótese é importante para dar plausibilidade à proposta de abordagem integrada da produção e da percepção do ritmo apresentada na seção 1.4. Os resultados dos dois experimentos, sobretudo os do experimento de percepção II, constituem evidência preliminar no sentido de corroborar a hipótese em questão ao mostrar que uma fronteira prosódica, marcada pela ocorrência de um acento frasal, afeta o comportamento dos ouvintes na realização da tarefa de monitoramento da ocorrência de cliques distribuídos ao longo de dois grupops acentuais. A seção 5.3 sugere a realização de variações do experimento de percepção II em busca de mais confirmação da hipótese do uso das unidades V-V como unidade básica da percepção do ritmo pelos falantes.

Uma das contribuições relevantes e originais que os experimentos de percepção realizados na tese trouxeram para a investigação experimental da percepção do ritmo é que a elaboração da tarefa experimental foi motivada pela interação com a pesquisa realizada no campo da produção. A tarefa de monitorar cliques alinhados às proximidades dos *onsets* vocálicos foi proposta exatamante porque no modelo dinâmico do ritmo a produção de intervalos V-V é o primitivo relevante modelado. Os dados de percepção obtidos podem, portanto, ser colocados em relação com os dados da produção de forma mais natural. Essa estreita relação que é possível estabelecer entre a produção e a percepção que é possibilitada pelo uso de dispositivos teórico-metodológicos da teoria dos sistema dinâmicos como os osciladores acoplados parece ser a chave que poderá levar a bom termo em etapas posteriores o empreendimento da construção de um modelo comum de produção e percepção do ritmo da fala.

Apêndice A

Algoritmos para análise da duração

Neste apêndice serão descritos em maiores detalhes os algoritmos utilizados para as análises dos dados de duração apresentadas na tese. Os algoritmos são idênticos ou versões apenas ligeiramente modificadas daqueles apresentados e implementados no script do programa Praat `SGDetector` em Barbosa (2006, pp. 169-170 e apêndice A.2).

Os algoritmos descritos neste apêndice permitem delimitar de forma automática os grupos acentuais de um trecho arbitrário de fala em análise a partir da duração acústica bruta de segmentos agrupados em unidades V-V. Essa delimitação é o produto final da aplicação de três procedimentos, descritos em seções separadas: normalização da duração bruta por *z-score* estendido (seção A.1), suavização do contorno de duração normalizado por meio da aplicação de uma função de média móvel de cinco pontos (seção A.2) e detecção de picos no contorno de duração suavizada (seção A.3). A seção A.4 traz um exemplo da aplicação dos algoritmos a título de exemplificação.

Os algoritmos, tomados em conjunto, fornecem uma definição de trabalho para a proeminência prosódica que tem a vantagem de ser explícita e replicável.

A.1 Normalização da duração por z-score

A.1.1 Justificativa

A adoção do procedimento de normalização da duração por z-score no contexto de experimentos prosódicos é justificada em razão da natureza relativa das proeminências. Como o interesse é comparar a força relativa de unidades contíguas, se a duração das unidades for comparada sem nenhum tipo de transformação prévia, a comparação pode ser distorcida pela influência da duração intrínseca, mascarando a influência do contexto prosódico sobre a duração dos fones. Esta distorção se manifesta da seguinte maneira: considere a comparação hipotética entre diversas realizações dos segmentos [ra] e [ta]. Mesmo no caso em que as realizações da primeira sejam consideravelmente alongadas em relação à média da classe em função de estarem num contexto prosódico que favoreça seu alongamento, elas serão possivelmente menores do que as sílabas [ta], mesmo se estas estiverem num contexto prosódico fraco, uma vez que a duração do fone [ɾ] é intrinsecamente menor que a de [t].

O procedimento pode ser aplicado aos fones quer estejam agrupados em unidades comparáveis com a sílaba fonológica quer em unidades V-V.

A.1.2 Implementação

A equação A.1 implementa a normalização por *z-score*. Cada unidade a ser normalizada é composta por um número n de fones e a variável *dur* representa a soma total das durações dos fones que compõem a unidade. As variáveis μ_i e var_i indicam respectivamente a média e a variância de cada um dos n fones componentes a serem somadas.

$$z = \frac{dur - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} var_i}} \tag{A.1}$$

Os valores de referência para μ e *var* de cada fone foram tomados de um *corpus* de logatomas contento cada combinação entre fones fonotaticamente possível na língua, gravado para a montagem do banco de unidades do sistema de síntese concatenativa de fala *Aiuruetê* (Barbosa et al. 1999). A lista contendo os valores de referência por ser encontrada em (Barbosa 2006, p. 489).

A.2 Suavização do contorno de duração

O algoritmo de suavização toma como entrada a sequência de unidades previamente suavizas (z_1, z_2, ..., z_{n-1}, z_n) e produz como resultado a sequência suavizada (sm_1, sm_2, ..., sm_{n-1}, sm_n).

A implementação da suavização é dada pela equação A.2

$$sm_i = \frac{5 \cdot z_i + 3 \cdot z_{i-1} + 3 \cdot z_{i+1} + z_{i-2} + z_{i+2}}{13} \tag{A.2}$$

A função geral de suavização não pode ser aplicada am algumas posições na sequência de unidades suavizadas. É o caso as duas primeiras unidades ($i = 1$ e $i = 2$) e as duas últimas unidades ($i = n - 1$ e $i = n$). Para elas aplica-se a equação A.3.

$$sm_i = \begin{cases} (2 \cdot z_i + z_{i+1})/3 & \text{se } i = 1, \\ (2 \cdot z_i + z_{i-1})/3 & \text{se } i = n, \\ (3 \cdot z_i + z_{i-1} + z_{i+1})/5 & \text{se } i = 2 \text{ ou } i = n - 1. \end{cases} \tag{A.3}$$

A.3 Detecção dos picos de duração

A entrada desta etapa é a sequência de durações suavizadas (sm_1, sm_2, ..., sm_{n-1}, sm_n). A próxima etapa é a detecção de "picos" no contorno formado pela sequência de durações suavizadas. Um "pico" é um ponto de máximo no contorno. Cada unidade sm_i será testada para a condição de pico e receberá o valor convencional 1 se for um máximo e 0 caso não seja. A saída do algoritmo é uma sequência de 0s e 1s. No quadro de trabalho que estamos adotando, uma unidade sm_i que recebe valor 1 marca a fronteira (*boundary*) de um grupo acentual.

A equação geral A.4 implementa a detecção de picos para:

$$bound_i = \begin{cases} 1 & \text{se } sm_{i-1} \leq sm_i > sm_{i+1}, \\ 0 & \text{nos demais casos.} \end{cases} \tag{A.4}$$

A função geral não se aplica ao primeiro e último termos da sequência, para as quais valem a equação A.5.

$$bound_i = \begin{cases} 1 & \text{se } sm_i > sm_{i+1} \text{ para } i = 1, \\ 1 & \text{se } sm_i \geq sm_{i-1} \text{ para } i = n, \\ 0 & \text{nos demais casos.} \end{cases} \tag{A.5}$$

A.4 Exemplo de aplicação

A presente seção mostra de forma concreta a aplicação dos algoritmos descritos nas seções anteriores a uma frase-veículo extraída do *corpus* do experimento de produção I, explicitando as entradas e saídas de cada uma das etapas.

O material escolhido para a exemplificação é o trecho em destaque de uma das repetições da frase-veículo "*Os edessenos budistas* parecem menores". No trecho a ser analisado são delimitadas oito unidades V-V (em notação ortográfica):

/os/$_1$, /ed/$_2$, /ess/$_3$, /en/$_4$, /os#b/$_5$, /ud/$_6$, /ist/$_7$ e /as#p/$_8$.

A entrada para o primeiro algoritmo, a normalização da duração (seção A.1), é uma lista, que será chamada de *ms*, obtida através da marcação das fronteiras das unidades V-V feita pelo analista, contendo a duração bruta (em milissegundos) de cada uma das oito unidades V-V ao longo do trecho a ser analisado, na ordem apresentada no parágrafo anterior. No caso do exemplo em discussão, a lista *ms* é igual a (133, 122, 176, 135, 182, 183, 207, 224). A saída produzida pela etapa de normalização é uma lista, chamada aqui de z, contendo os oito valores normalizados correspondentes a cada elemento da lista *ms*. A lista z do trecho analisado é

$$(-1.29, -1.67, -0.98, -1.02, -1.4, -1.43, -0.84, -1.09).$$

A lista z, é ao mesmo tempo saída da etapa de normalização e entrada do segundo algoritmo, a suavização da duração normalizada (seção A.2). O segundo algoritmo produz uma lista, *sm*, a partir de z. Cada elemento de *sm* é o correspondente suavizado do elemento de mesmo índice em z. No caso do exemplo discutido aqui *sm* é igual a

$$(-1.42, -1.46, -1.2, -1.18, -1.24, -1.23, -1.01, -1.01).$$

A lista *sm*, saída da etapa de normalização, funciona como entrada para a última etapa, a detecção de picos locais no contorno de duração suavizada (seção A.3). A saída do algoritmo de detecção é uma lista, b, que indica, de forma binária, se cada elemento de *sm* é ou não um pico local no contorno definido pelos valores de *sm*. Barbosa (2006) define os picos detectados a partir deste procedimento como indicadores de fronteira prosódica ao nível da produção. A lista b para o exemplo analisado é

(1, 0, 0, 1, 0, 0, 1, 0).

Uma representação alternativa mais compacta para b é uma lista que omite os elementos nulos e mostra os índices das unidades V-V que correspondem a fronteiras prosódicas ou picos na lista *sm* separados por uma barra vertical (|). No caso do exemplo analisado nesta seção, essa representação compacta seria:

1|4|7.

Há ainda a possibilidade de integrar a informação contida em b na representação ortográfica do trecho analisado. Observe-se que a colocação das barras verticais (|), indicativas da fronteira, respeita as fronteiras sílábicas:

|Os edesse|nos budis|tas

A figura A.1 mostra as transformações sofridas pelo contorno duracional bruto da frase usada como exemplo à medida que ele passa pelas etapas de normalização e suavização e a localização dos picos no contorno suavizado.

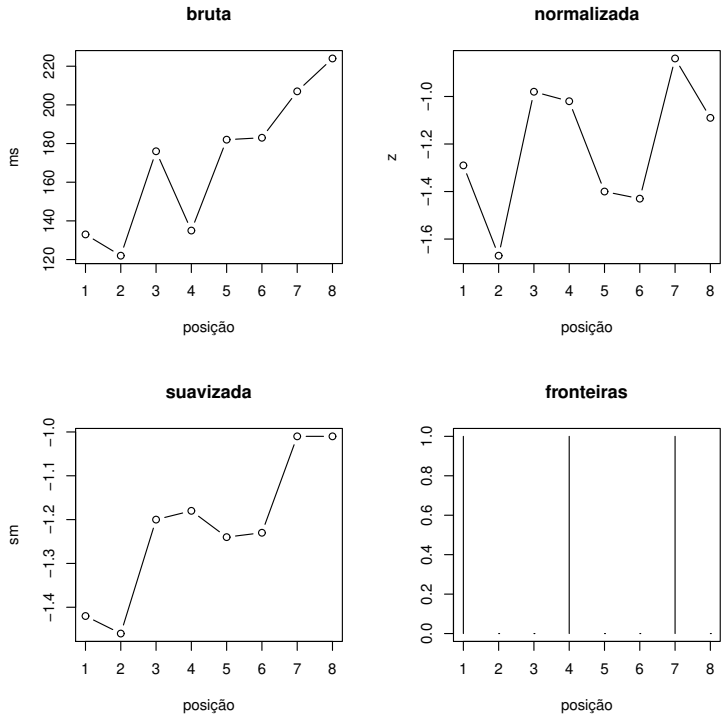


Figura A.1: Contornos duracionais definidos pela duração bruta (lista *ms*), normalizada (z), suavizada (sm) e localização dos picos da duração suavizada (b) na frase de exemplo.

Apêndice B

Algoritmo para análise de f_0

B.1 Justificativa

Uma das abordagens mais empregadas para análise da frequência fundamental (f_0) como possível correlato da acentuação que é possível encontrar na literatura consiste em reduzir toda a variabilidade do contorno de f_0 a apenas um valor de f_0 para cada vogal no trecho a ser analisado. Para exemplificar, veja-se a figura abaixo.

Na figura B.1 vê-se o contorno de f_0 calculado pelo programa Praat representado pelos círculos azuis abertos[1]. Cada um dos quadrados vermelhos cheios representa o valor de f_0 (obtido por interpolação) medido no ponto médio do intervalo de duração acústica de cada uma das vogais do trecho. O contorno delimitado pela linha contínua que liga os quadrados vermelhos é a representação da curva original de f_0 do trecho do enunciado em questão gerado pela abordagem da estilização por pontos quando o número de pontos por segmento vocálico é igual a um.

A comparação dos dois contornos mostra que, embora bastante simplificada, a estilização usando um ponto por vogal é capaz de, em alguma medida, resumir algumas características da curva original. Ele mostra uma subida quase linear do valor de f_0 até a terceira vogal do trecho, seguida de uma queda na vogal tônica da palavra gonococo e, finalmente, uma nova subida na vogal pós-tônica. Esse cenário é em certa medida razoável, pois de fato há uma variação considerável de f_0 ocorrendo durante as duas primeiras vogais que culmina em torno da terceira vogal. O ponto relativamente mais baixo na vogal correspondente à sílaba tônica seguido de um ponto mais alto na pós-tônica também faz sentido. Uma das vantagens dessa estilização é que é, fazendo-se a média do valor de cada ponto de uma série de repetições da mesma frase, é viável aplicar técnicas de estatística inferencial para comparar os valores de f_0 ponto a ponto ao longo da frase. Com um número maior de pontos a aplicação de técnicas de inferência não é em tese impossível, mas tende a se tornar impraticável, pois o número de comparações pareadas começaria a subir muito rapidamente.

A simplificação drástica da variação complexa da curva de f_0 representada pela técnica exemplificada acima introduz, apesar de sua parcial adequação, impropriedades não desprezíveis

[1] Cada ponto é uma estimativa da f_0 em um intervalo de 10 ms, note-se a falta de pontos nos intervalos desvozeados, correspondentes aos fones [k].

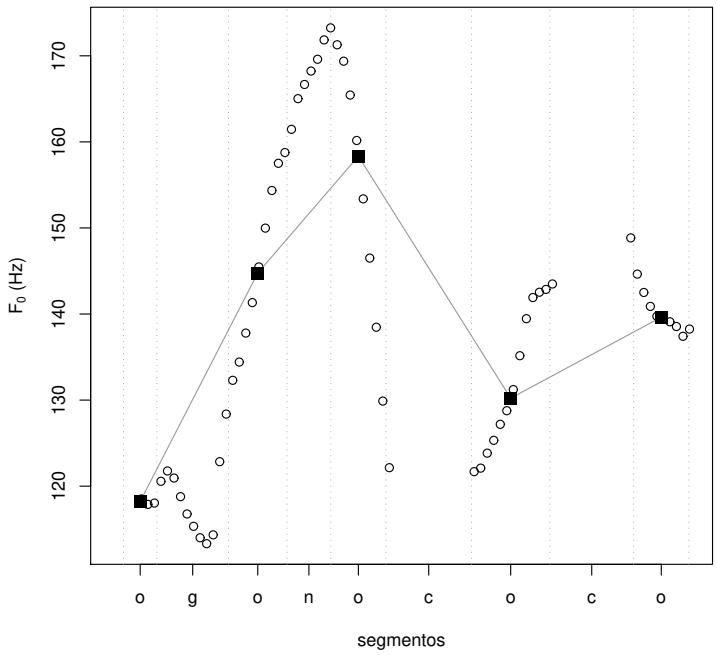


Figura B.1: Contorno de f_0 do grupo acentual "O gonococo", na frase *O gonococo parece menor hoje*. Os círculos mostram os valores de f_0 calculados pelo programa Praat (janela de análise de 10 ms). Os quadrados ligados pelo traço contínuo mostram o valor interpolado de f_0 no ponto médio da duração acústica das vogais. As linhas verticais pontilhadas delimitam os intervalos de duração acústica de cada segmento.

no contorno estilizado. Uma dessas impropriedades é o fato da estilização não refletir, no exemplo da figura B.1, o fato da variação observada na segunda vogal ser quase especularmente simétrica à variação da terceira vogal, isto é, a amplitude de variação positiva observada na segunda vogal é quase igual à amplitude negativa de variação observada na terceira[2]. A estilização também deixa escapar que o valor baixo do ponto na quarta vogal é resultado da continuação da queda de f_0 observada na terceira vogal e que, apesar dessa inércia, nessa mesma vogal tem início um movimento ascendente na curva de f_0. O valor relativamente mais alto do ponto na quinta vogal da figura B.1 é a consequência, portanto, de um movimento cujo início ocorreu na vogal precedente.

[2]Na figura B.1, a diferença entre os pontos referentes à segunda e à terceira vogais é de pouco mais de 13 Hz. A comparação estatística das médias para estes mesmos dois pontos, obtidas em 50 repetições dessa mesma frase mostra que não há diferença significativa entre os dois valores. Uma vez que o movimento é quase especularmente simétrico, é de se esperar que no ponto médio os valores de f_0 sejam parecidos. O que a estilização não é capaz de captar é o fato de que na segunda vogal a f_0 passa pelo ponto médio com trajetória ascendente e na terceira passar aproximadamente pelo mesmo valor, mas com trajetória descendente. Isso quer dizer que a estilização é estática, isto é, não é capaz de diferenciar as duas posições em ternos dos movimentos que ocorrem em cada vogal.

Com o objetivo de tentar superar as inadequações e insuficiências do tipo deste tipo de estilização desenvolveu-se um algoritmo para a análise de f_0 que tirasse proveito de informações dinâmicas da curva entoacional. Os elementos que essa abordagem busca identificar são movimentos significativos da curva de f_0 que ajudem a entender como a interação entre os mecanismos de produção da fala e as funções comunicativas se combinam e dão origem à complexa variabilidade observada nos contornos entoacionais particulares.

Os movimentos a que faz referência o parágrafo anterior são os pontos máximos e mínimos (locais e globais) da curva de f_0 e da curva da derivada de f_0, que pode ser entendida como a taxa de mudança (velocidade) de f_0 ao longo do tempo. A idéia é tentar buscar correlações que possam vir a sugerir que esses pontos singulares, como picos e vales no contorno entoacional, estão relacionados com intenções linguísticas dos falantes, como sinalizar a modalidade de um enunciado, ou uma proeminência, por exemplo.

B.2 Implementação

O algoritmo para encontrar os pontos extremos dos contornos de f_0 e f_0' é composto pelas seguintes etapas:

1. A curva de f_0 do arquivo de som é extraída. Valores espúrios podem ser opcionalmente removidos manualmente após inspeção visual da curva;

2. A curva de f_0 é suavizada por meio da função *Smooth* do programa Praat. Os melhores resultados em termos da remoção de oscilações micromelódicas são obtidos quando a função de suavização é executada usando uma largura de banda entre 3 e 5 Hz;

3. Aplica-se uma função que interpola, usando uma função quadrática, os trechos desvozeados da curva suavizada. A sequência $F = (p_1,\ p_2,\ \ldots,\ p_i,\ \ldots,\ p_{n-1},\ p_n)$, em que cada elemento p_i corresponde a um valor de f_0 da curva suavizada e interpolada, é produzida. Supõe-se que a curva representada por este conjunto de pontos represente de forma aproximada o componente macromelódico (no sentido em que essa palavra é usada por Hirst em [41]) do enunciado;

4. Calcula-se a derivada da sequência obtida no passo anterior por meio da diferenciação discreta, de forma a obter a sequência $F' = (q_1,\ q_2,\ \ldots,\ q_j,\ \ldots,\ q_{m-1},\ q_m)$, com $m = n-1$. Cada ponto q_j é obtido pela operação de diferenciação discreta, definida por $q_j = p_i - p_{i-1}$, para $i > 1$ e $j = i - 1$[3];

[3] A rigor, o segundo termo da operação de diferenciação deveria ser $(p_i - p_{i-1})/\Delta t$. O incremento de tempo Δt foi omitido neste caso por ser fixo (por isso a operação de diferenciação é classificada nesse caso como discreta). Aqui, o incremento Δt corresponde ao quadro de análise, isto é, o intervalo de tempo que define o período de amostragem usado pelo algoritmo de extração de f_0. Na atual versão do *script* que implementa o algoritmo descrito nesta seção, o quadro tem um valor de 10 ms, o que significa que os elementos das sequências de pontos gerados pelo algoritmo estão uniformemente separados por intervalos de 10 ms. Como a divisão por 10 foi omitida, o valor de f_0' está expresso em Hertz por quadro (Hz/dx, dx é a notação convencional para o incremento Δt). Alternativamente, os valores de f_0' podem ser expressos em dHz/ms (1 dHz = 1 Hz/10).

5. Os pontos extremos (máximos e mínimos) das curvas de f_0 e f_0' são identificados. Cada máximo é um elemento das sequências F ou F' que satifaz a condição $p_{i-1} < p_i > p_{i+1}$. Cada mínimo é um elemento de F ou F' que satifaz a condição $p_{i-1} > p_i < p_{i+1}$;

6. Os pontos de F identificados como máximos recebem o rótulo H(igh) e recebem o rótulo R(ising) quando se trata dos máximos de F'. Os pontos de F identificados como mínimos recebem o rótulo L(ow) e os de F' são etiquetados F(alling). Cada elemento extremo (conjunto dos máximos e mínimos) é pareado com o ponto t_i que representa a coordenada temporal daquele ponto extremo.

A figura B.2 mostra o resultado da aplicação do *script* no mesmo arquivo de som ilustrado na figura B.1. Na parte de cima da figura pode-se ver o contorno de f_0 suavizado (por meio da função *Smooth* do Praat). Os traços verticais verdes mostram os pontos de máximo do contorno e os vermelhos os pontos de mínimo. O mesmo vale para o contorno da velocidade de f_0 mostrado na parte de baixo da figura. A segmentação nos retângulos azuis entre os dois contornos mostra o alinhamento de picos e vales com os segmentos, agrupados em unidades V-V.

Na implementação do algoritmo descrito acima como um *script* do Praat, os pontos extremos são marcados em uma camada *PointTier* em um *TextGrid*, que contém também a segmentação dos fones agrupados em unidades V-V[4]. A sobreposição dos pontos extremos com a segmentação das unidades V-V permite extrair pelo menos dois tipos de informação:

1. o número de pontos extremos num determinado trecho da frase analisada;

2. o valores numérico associado ao ponto extremo. No caso dos extremos do tipo H e L o valor é dado em Hz e nos extremos do tipo R e F o valor corresponde à velocidade de mudança de f_0 dada em Hz/dx.

Além disso, é possível, ainda, quantificar o alinhamento de cada ponto extremo em relação à unidade segmentada em que o ponto se encontra. A fase (ϕ) é obtida da seguinte forma (veja também figura B.3):

1. Determina-se, para o extremo E_{ij}, associado à unidade V-V$_i$, o intervalo de tempo entre o instante de início da unidade V-V e o instante de tempo em que se encontra o ponto extremo em questão (Δt_{ij}).

2. Divide-se Δt_{ij} pelo valor da duração da unidade V-V$_i$ a que o ponto extremo está associado e multiplica-se o resultado por 100 para que os valores fiquem entre 0 e 100%. A figura B.3 mostra duas situações especiais. A primeira delas se dá quando um extremo está associado à unidade V-V$_1$, isto é, à primeira unidade da segmentação. Se esse extremo, a que faremos referência por E_{1j}, estiver temporalmente localizado antes do início de V-V$_1$, então o valor da fase para este extremo, ϕ_{1j}, será negativo. O outro caso especial ocorre quando um extremo está associado à V-V$_n$, isto é, à última da segmentação. Se um extremo E_{nj} estiver estiver localizado temporalmente depois do final de V-V$_n$, então ϕ_{nj} terá um valor maior do que 100%.

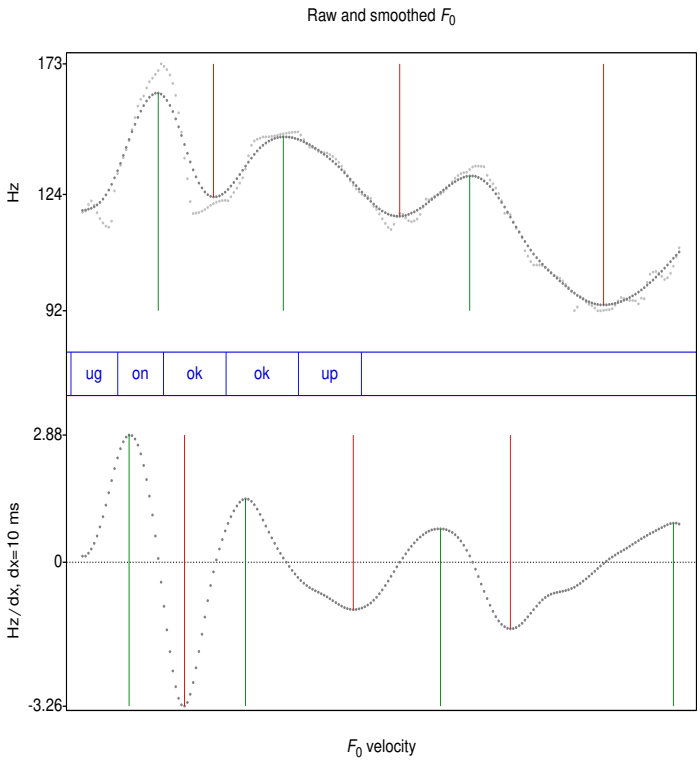


Figura B.2: Resultado da aplicação do *script* que detecta os pontos extremos. Na parte superior os círculos cinza claro correspondem aos valores brutos e os cinza escuro aos valores suavizados de f_0. Na parte inferior encontra-se o contorno da velocidade de f_0 obtido a partir da diferenciação do contorno suavizado de f_0 visto na parte de cima da figura. Nos dois contornos, os traços verticais verdes marcam os máximos e os traços verticais vermelhos os mínimos. Entre os dois contornos a segmentação em unidades V-V mostra o alinhamento entre os pontos extremos e as fronteiras das unidades V-V.

A fórmula B.1 mostra como é feito o cálculo de fase ϕ_{ij} para um extremo E_{ij} associado a uma unidade V-V$_n$ da segmentação.

$$\phi_{ij} = \frac{\Delta t_{ij}}{dur_{V-V_i}} \times 100 \tag{B.1}$$

Para exemplificar as definições dadas acima, observe-se a figura B.2. A fase do primeiro máximo de f_0 será calculada relativamente a seu alinhamento com a segunda unidade V-V, isto é, |on|. É o mesmo caso do primeiro máximo de f_0', cuja fase será calculada relativamente à mesma unidade V-V. Já a fase do primeiro mínimo de f_0 será calculada relativamente à terceira unidade V-V, isto é, a primeira ocorrência de |ok|.

Uma característica importante da curva de f_0 mostrada na figura B.2 que a abordagem por pontos extremos captura e que a abordagem descrita no início do anexo não dá conta é a quase simetria existente entre os movimentos de f_0 que ocorrem alinhados às unidades |on| e |ok|.

[4] O usuário pode fornecer uma outra segmentação qualquer.

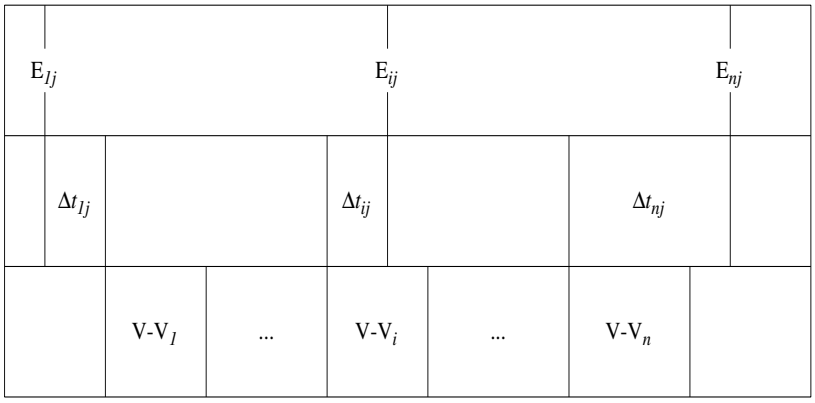


Figura B.3: Representação esquemática do alinhamento entre as marcações de duração das unidades V-V e pontos extremos.

Essa simetria se manifesta no contorno de f_0'. O pico positivo de f_0' (R), alinhado com o final da unidade |on| tem um valor muito semelhante com o valor absoluto do primeiro pico negativo de f_0' (F), alinhado aproximadamente com a metade da primeira unidade |ok|. É possível verificar visualmente que quase todo o aumento de f_0 que ocorre em |on| é perdido em |ok|.

Apêndice C

Experimento de produção II

C.1 Lista das frases-veículo do *corpus*

Segue abaixo a lista completa das frases-veículo do *corpus* do experimento. Abaixo de cada palavra-alvo estão listadas quatro frases. Nos três primeiros itens a palavra-alvo está, respectivamente, em posição inicial, medial e final na frase (ver seção 4.2). No quarto item a palavra-alvo tem diferentes *status* referencial em cada uma das orações (ver serção 4.3).

- deputado
  - O deputado passou por aqui hoje.
  - Pedro viu o deputado passar por aqui.
  - Paulo viu o deputado. Ele passou por aqui hoje.
  - Um deputado foi depor no plenário Câmara Federal.
    O deputado foi evasivo em suas declarações.
- jornalista
  - O jornalista passou por aqui hoje.
  - Marcos viu o jornalista passar por aqui.
  - João viu o jornalista. Ele passou por aqui hoje.
  - Um jornalista sentou para escrever sua coluna no último minuto.
    O jornalista estava sem idéias naquele dia.
- delegado
  - O delegado passou por aqui hoje.
  - Rita viu o delegado passar por aqui.
  - Rosa viu o delegado. Ele passou por aqui hoje.
  - Um delegado comandou uma operação arriscada de busca na fronteira.
    O delegado prendeu uma quadrilha perigosa.

- jogador
  - O jogador passou por aqui hoje.
  - Rubens viu o jogador passar por aqui.
  - Dora viu o jogador. Ele passou por aqui hoje.
  - Um jogador foi vendido para um time estrangeiro por muito dinheiro. O jogador deixou saudades na torcida.
- senador
  - O senador passou por aqui hoje.
  - Lucas viu o senador passar por aqui.
  - Hugo viu o senador. Ele passou por aqui hoje.
  - Um senador foi à tribuna e fez um discurso inflamado. O senador tinha uma opinião forte.
- militar
  - O militar passou por aqui hoje.
  - Ana viu o militar passar por aqui.
  - Bia viu o militar. Ele passou por aqui hoje.
  - Um militar foi achado na selva depois de pedir socorro. O militar ficou perdido na mata por dias.
- americano
  - Os americanos passaram por aqui hoje.
  - Carlos viu o americano passar por aqui.
  - Alda viu o americano. Ele passou por aqui hoje.
  - Uns americanos fecharam um negócio lucrativo no Brasil. Os americanos vão plantar cana-de-açúcar em São Paulo.
- departamento
  - O departamento fechou mais cedo hoje.
  - Celso viu o departamento fechar cedo.
  - Jonas foi ao departamento. Ele fechou cedo hoje.
  - Um departamento será criado em breve pelo prefeito. O departamento deve cuidar dos assuntos financeiros.
- secretaria
  - A secretaria fechou mais cedo hoje.

  - Chico viu a secretaria fechar cedo.
  - Ivo foi à secretaria. Ela fechou cedo hoje.
  - Uma secretaria será criada em breve pelo prefeito.
    A secretaria deve cuidar de assuntos financeiros.

- governador
  - O governador passou por aqui hoje.
  - Gilda viu o governador por aqui.
  - Ivan viu o governador. Ele parecia diferente hoje.
  - Um governador sofreu críticas duras de toda a imprensa.
    O governador botou parentes em cargos públicos.

- consumidor
  - O consumidor passou por aqui hoje.
  - Jorge viu o consumidor por aqui.
  - Roque viu o consumidor. Ele parecia diferente hoje.
  - Um consumidor fez uma queixa contra uma loja que atrasou a entrega.
    O consumidor tinha se sentido lesado pela loja.

- procurador
  - O procurador passou por aqui hoje.
  - Luiz viu o procurador passar por aqui.
  - Raquel viu o procurador. Ele parecia diferente hoje.
  - Um procurador foi ameaçado pelos criminosos.
    O procurador fez denúncias graves.

- universidade
  - As universidades fazem prova hoje.
  - Rute viu as universidade crescerem aqui.
  - Bruno vai às universidades. Hoje é dia de prova.
  - Uma possibilidade apareceu de repente na vida de José.
    A possibilidade partiu da indicação de um amigo.

- oportunidade
  - As oportunidades somem no país hoje.
  - Artur viu as oportunidades sumirem no país.
  - Sara viu as oportunidades. Elas somem no país.

  - Umas oportunidades surgiram na agência de empregos.
    As oportunidades apareceram por conta do Natal.

- possibilidade

  - A possibilidade nasce a cada dia.
  - Sandro viu a possibilidade crescer a cada dia.
  - Davi viu a possibilidade. Ela cresce a cada dia.
  - Uma possibilidade apareceu de repente na vida de José.
    A possibilidade partiu da indicação de um amigo.

- privatização

  - A privatização cresce no país hoje.
  - A bolsa vê a privatização crescer no país.
  - A bolsa vê a privatização. Ela cresce no país hoje.
  - Uma privatização causou medo entre os empregados.
    A privatização resultaria em possíveis demissões.

- sofisticação

  - A sofisticação passou na moda atual.
  - Elza não vê sofisticação na moda atual.
  - Leila quer sofisticação. Ela acabou na moda atual.
  - Uma sofisticação surpreendeu a platéia no desfile.
    A sofisticação parecia ter acabado na moda atual.

- peregrinação

  - A peregrinação saiu daqui cedo.
  - O padre viu a peregrinação passar aqui.
  - O padre viu a peregrinação. Ela passou por aqui.
  - Uma peregrinação reuniu muitas pessoas em volta da igreja.
    A peregrinação durou o dia todo.

Referências Bibliográficas

[1] Albano, Eleonora C.: *O gesto e suas bordas: esboço de Fonologia acústico-articulatória do português brasileiro*. Editora Mercado de Letras, Campinas, 2001.

[2] Ali, Manuel Said: *Difficuldades da Lingua Portugueza: Estudos e Observações*. Laemmert & c., Rio de Janeiro, 1908 [1985].

[3] Arantes, Pablo e Plinio Almeida Barbosa: *Acentuação secundária em Português Brasileiro à luz de modelo dinâmico do ritmo: um estudo piloto*. In *Anais do I Congresso Internacional de Fonética e Fonologia/VII Congresso Nacional de Fonética e Fonologia*, Belo Horizonte–MG, 2002.

[4] Barbosa, P. A., P. Arantes e L. S. Silveira: *Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule*. In *Proceedings of the Speech Prosody 2004 Conference*, páginas 49–52, Nara, Japan, 2004.

[5] Barbosa, Plinio A.: *Elementos para uma tipologia do ritmo (lingüístico) à luz de um modelo de osciladores acoplados*. In Cognito, 2(1):31–58, 2004.

[6] Barbosa, Plinio A.: *Incursões em torno do ritmo da fala*. Pontes, Campinas, 2006.

[7] Barbosa, Plinio A.: *From syntax to acoustic duration: a dynamical model of speech rhythm production*. Speech Communication, 49(1–2):75–96, 2007.

[8] Barbosa, Plinio Almeida: *Caractérisation et génération automatique de la structuration rythmique du français*. Tese de Doutoramento, Institut de la Communication Parlée, Université Stendhal, Grenoble, 1994.

[9] Barbosa, Plínio Almeida: *At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration*. In *Proceedings of the 1st ESCA Tutorial Research Workshop on Speech Production Modeling and 4th Speech Production Seminar*, páginas 85–88, Autrans, France, 1996.

[10] Barbosa, Plínio Almeida: *Generating duration from a cognitively plausible model of rhythm production*. In Association, European Speech Communication (editor): *Proceedings of the 7th European Conference on Speech Communication and Technology*, volume 2, páginas 967–970, Ålborg, Dinamarca, 2001.

[11] Barbosa, Plinio Almeida: *Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production.* Caderno de Estudos Lingüísticos, 43:71–92, 2002.

[12] Barbosa, Plínio Almeida: *Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production.* In *Proceedings of the Speech Prosody 2002 Conference*, páginas 163–166, Aix-en-Provence, France, 2002.

[13] Barbosa, Plínio Almeida e Sandra Madureira: *Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese.* In *Proceedings of the XIVth International Congress of Phonetic Sciences*, volume 1, páginas 297–300, San Francisco, USA, 1999.

[14] Barbosa, Plínio, et. al.: *Aiuruetê: a high quality concatenative text-to-speech system for Brazilian Portuguese.* In ECA (editor): *Proc. of the 6th European Conference on Speech Communication and Technology*, volume 5, páginas 2059–2062, 1999.

[15] Bertinetto, Pier Marco: *L'accento secondario nella fonologia italiana – Analise teorica e sperimentale.* In Simone, Raffaele, Ugo Vignuzzi e Giulianella Ruggiero (editores): *Studi de Fonetica e Fonologia*, páginas 189–235. Bulzoni, Roma, 1976.

[16] Bertinetto, Pier Marco: *Strutture prosodiche dell'italiano*, capítulo Accento secondario, páginas 91–114. Presso L'Accademia della Crusca, Firenze, 1981.

[17] Box, G. E. P. e D. R. Cox: *An analysis of transformations.* Journal of the Royal Statistical Society, Series B, 26(2):211–252, 1964.

[18] Browman, Catherine P. e Louis Goldstein: *Towards an articulatory phonology.* Phonology Yearbook, 3:219–252, 1986.

[19] Browman, Catherine P. e Louis Goldstein: *Articulatory gestures as phonological units.* Phonology, 6:201–251, 1989.

[20] Browman, Catherine P. e Louis Goldstein: *Tiers in articulatory phonology, with some implications for casual speech.* In Kingston, John e Mary E. Beckman (editores): *Papers in laboratory phonology I: between the grammar and the physics of speech*, páginas 341–376. Cambridge University Press, Cambridge, MA, 1990.

[21] Brown, G.: *Prosodic structure and the Given/New distinction.* In *Prosody: Models and Measurements.* Springer-Verlag, Berlin, 1983.

[22] Bybee, Joan L.: *Usage-based Phonology.* In Darnell, Michael *et al.* [29], páginas 211–242.

[23] Bybee, Joan L.: *Phonology and Language Use.* Cambridge University Press, Cambridge, MA, 2001.

[24] Castelo, Adelina: *A Proeminência Secundária no Português Europeu.* Tese de Mestrado, Universidade de Lisboa, Lisboa, 2004.

[25] Collischonn, Gisella: *Acento Secundário em Português*. Letras de Hoje, 29:43–53, 1994.

[26] Connine, Cynthia e Debra Titone: *Phoneme Monitoring*. Language and Cognitive Processes, 11(6):635–645, 1996.

[27] Cutler, Anne: *Speaking for listening*. In Allport, Alan, Donald Mackay, Wolfgang Prinz e Eckart Scheerer (editores): *Language Perception and Production*, páginas 23–40. Academic Press, New York, 1987.

[28] Cutler, Anne e Donald J. Foss: *On the role of sentence stress in sentence processing*. Language and Speech, 20(1):1–10, 1975.

[29] Darnell, Michael, Edith Moravcsik, Frederick Newmeyer, Michael Noovan e Kathleen Wheatley (editores): *Functionalism and Formalism in Linguistics*, volume I: General Papers. John Benjamins Publishing Company, Amsterdam, 1999.

[30] de Jong, K., M. Beckman e J. Edwards: *The interplay between prosodic structure and coarticulation*. Language and Speech, 26:197–212, 1993.

[31] Dechaine, R. M.: *Stress and weight gain*, 1990. Manuscrito.

[32] Dogil, Grzegorz: *Phonetic correlates of rhythmic stress in Spanish*. In van der Hulst, Harry (editor): *Word Prosodic Systems in the Languages of Europe*, páginas 304–311. Mouton de Gruyter, Berlin, 1999.

[33] Forster, K. I. e J. C. Forster: *DMDX: A Windows display program with millisecond accuracy*. Behavior Research Methods, Instruments, & Computers, 35(1):116–124, 2003.

[34] Fowler, C. e J. Housum: *Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction*. Journal of Memory and Language, (26):489–504, 1987.

[35] Francis, W. Nelson e Henry Kučera: *Frequency analysis of English usage*. Houghton Mifflin, Boston, 1982.

[36] Gama Rossi, Aglael Juliana: *Qual é a natureza do acento secundário no português brasileiro?* Cadernos Centro Universitário São Camilo, 4:77–92, 1998.

[37] Garde, Paul: *L'Accent*. PUF, Paris, 1968.

[38] Gonçalves Viana, Aniceto R.: *Exposição da pronúncia normal portuguesa para uso de nacionais e estrangeiros*. In *Estudos de fonética portuguesa*, páginas 153–259. Imprensa Nacional/Casa da Moeda, Lisboa, 1973[1892].

[39] Gundel, Jeanette K., Nancy Hedberg e Ron Zacharski: *Cognitive Status and the Form of Referring Expressions in Discourse*. Language, 69(2):274–307, 1993.

[40] Hammond, Michael: *Lexical frequency and rhythm*. In Darnell, Michael *et al.* [29], páginas 329–358.

[41] Hirst, D. J.: *Form and function in the representation of speech prosody.* Speech Communication, 3–4(46):334–347, 2005.

[42] Hoskins, Stevens R.: *Secondary stress and stress clash resolution in French: an empirical investigation.* In Mazzola, Michael L. (editor): *Issues and theory in Romance Linguistics: selected papers from the linguistic symposium on romance languages XXIII April 1-4*, páginas 35–47. Georgetown University Press, Washington, DC, 1994.

[43] Juilland, Alphonse e Eugenio Chang-Rodriguez: *Frequency Dictionary of Spanish Words.* Mouton, The Hague, 1964.

[44] Kelso, J. A. S.: *Dynamic Patterns: The Self-Organization of Brain and Behavior.* MIT Press, Cambridge, MA, 1995.

[45] Kelso, J. A. Scott, Elliot L. Saltzman e Betty Tuller: *The dynamical perspective in speech production: data and theory.* Journal of Phonetics, 14:29–59, 1986.

[46] Lee, Seung Hwa: *Acento secundário do PB.* Letras de Hoje, 37:149–162, 2002.

[47] Lehiste, Ilse: *Suprasegmentals.* MIT Press, Cambridge, 1970.

[48] Liberman, Mark e Alan Prince: *On stress and linguistic rhythm.* Linguistic Inquiry, 8:249–336, 1977.

[49] Lindblom, Björn e Johan Sundberg: *Acoustical consequences of lip, tongue, jaw and larynx movement.* Jounal of the Acoustical Society of America, 50(4):1166–1179, 1971.

[50] Martin, James G.: *Aspects of rhythmic structure in speech perception.* In Evans, James R. e Manfred Clynes (editores): *Rhythm in psychological, linguistic and musical processes*, páginas 79–98. Charles C. Thomas Publiser, Springfield, IL, 1986.

[51] Martinet, André: *Éléments de linguistique générale.* Armand Colin, Paris, 1961.

[52] Massini, Gladis: *A duração no estudo do acento e ritmo do português.* Tese de Mestrado, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1991.

[53] The Mathworks: *Matlab 6.0 R12 Documentation*, 2001.

[54] Mattoso Câmara Jr., Joaquim: *Estrutura da Língua Portuguesa.* Vozes, Petrópolis, 1970.

[55] Mattoso Câmara Jr., Joaquim: *Problemas de lingüística descritiva.* Vozes, Petrópolis, 1970.

[56] Mattoso Câmara Jr., Joaquim: *História e Estrutura da Língua Portuguesa.* Padrão — Livraria e Editora LTDA., Rio de Janeiro, 1976.

[57] McAuley, J. Devin: *Perception of time as phase: towards an adaptive-oscillator model of rhythmic pattern processing.* Tese de Doutoramento, Indiana University, USA, Bloomington, IN, 1995.

[58] Meltzer, R. H., J. C. Martin, C. B. Mills, D. L. Imhoff e D. Zohar: *Reaction time to temporally-displaced phoneme targets in continuous speech.* Journal of Experimental Psychology: Human Perception and Performance, 2(2):277–290, 1976.

[59] Mitchell, D. C.: *An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading.* In Kieras, David E. e Marcel. A. Just (editores): *New Methods in Reading Comprehension Research.* Erlbaum, Hillsdale, N.J., 1984.

[60] Nascentes, Antenor: *O idioma nacional.* Livraria Acadêmica, Rio de Janeiro, 1960[1926].

[61] Nespor, Marina e Irene Vogel: *Prosodic Phonology.* Foris Publications, Dordrecht, 1986.

[62] Oliveira, Leonardo Couto Franco de: *Léxico, alofonia e percepção de fala na Fonologia Articulatória.* Tese de Mestrado, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2003.

[63] Prieto, Pilar e Jan van Santen: *Secondary stress in Spanish: some experimental evidence.* In Parodi, Claudia, Carlos Quicoli, Mario Saltarelli e Maria Luiza Zubizarreta (editores): *Aspects of Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages XXIV March 10-13*, páginas 337–356. Georgetown University Press, Washington, DC, 1996.

[64] Prince, Alan: *Relating to the grid.* Linguistic Inquiry, 14:19–100, 1983.

[65] Prince, Alan e Paul Smolensky: *Optimality Theory: Constraint Interaction in Generative Grammar.* Relatório Técnico 2, Rutgers University Center for Cognitive Science, 1993.

[66] Prince, Ellen F.: *Towards a tanoxnomy of given-new information.* páginas 223–56. Academic Press, 1981.

[67] Quené, Hugo e René Kager: *Prosodic sentence analysis without exhaustive parsing.* In van Heuven, Vincent J. e Louis C. W. Pols (editores): *Analysis and Synthesis of Speech*, páginas 115–130. Mouton de Gruyter, Berlin, 1993.

[68] Quené, Hugo e Robert Port: *Effects of timing regularity and metrical expectancy on spoken-word perception.* Phonetica, 62(1):1–13, 2005.

[69] Rocca, Iggy: *Secondary stress and metrical rhythm.* Phonology Yearbook, 3:341–370, 1986.

[70] Saltzman, Elliot L. e Kevin G. Munhall: *A Dynamical Approach to Gestural Patterning in Speech Production.* Ecological Psychology, 1(4):333–382, 1989.

[71] Sandalo, F., M. B. Abaurre, A. Mandel e C. Galves: *Secondary stress in two varieties of Portuguese and the Sotaq optimality based computer program.* Probus, 18:97–125, 2006.

[72] Scharf, G., I. Hertrich, I. Rocca e G. Dogil: *Articulatory correlates of secondary stress in polish and spanish.* In *Proceedings of The 7th International Conference of Phonetic Sciences*, volume 2, páginas 634–637, Stockholm, Sweden, 1995.

[73] Schütze, Carson T.: *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. University of Chicago Press, Chicago, IL, 1996.

[74] Strogatz, S. e I. Stewart: *Oscillateurs Couplés et Synchronisation Biologique*. Pour la Science, 196:40–46, 1994.

[75] Tranel, Bernard: *The sounds of French*. Cambridge University Press, Cambridge, 1986.

[76] Vallabha, Gautam K. e Betty Tuller: *Systematic errors in the formant analysis of steady-state vowels*. Speech Communication, 38:141–160, 2002.

[77] Vogel, Irene e Sergio Scalise: *Secondary stress in Italian*. Lingua, 58:213–242, 1982.

[78] Wolters, Maria: *Prosodic correlates of referent status*. In *Proceedings of the XVIth ICPHS*, San Francisco, CA, 1999.

[79] Zipf, George Kingsley: *The Psycho-Biology of Language*. The MIT Press, Cambridge, MA, 1965[1935].