

Working Papers
em Linguística



Fala, voz e expressividade

2018

v. 19

N. 2

Working Papers em Linguística, v. 19, n. 2, 2018

Centro de Comunicação e Expressão - CCE
Universidade Federal de Santa Catarina - UFSC
Florianópolis - SC - Brasil

Editora-chefe

Izete Lehmkuhl Coelho

Coeditores

Marco Antonio Martins
Izabel Christine Seara

Organização

Leticia Rebollo Couto
Izabel Christine Seara

Editoração

Ana Eltermann
Clóvis Alencar Butzge
Elaine Melo
Érica Marciano de Oliveira Zibetti
Gésyka Mafra
Hayla Thami
João Paulo Zarelli Rocha
Jonathan Ribeiro Farias de Moura
Lorena Cardoso
Raquel Gomes Chaves

Conselho Editorial

Adair Bonini, Universidade Federal de Santa Catarina, Brasil
Adja Balbino de Amorim Barbieri Durão, Universidade Federal de Santa Catarina, Brasil
Adriana Fischer, Centro Universitário de Brusque, Brasil
Aline Cacilda Koteski Emilio, Universidade Estadual de Ponta Grossa, Brasil
Ana Cláudia Souza, Universidade Federal de Santa Catarina, Brasil
Ana Paula Oliveira Santana, Universidade Federal de Santa Catarina, Brasil
André Berri, Universidade Federal de Santa Catarina, Brasil
Clarice Nadir von Borstel, Universidade Estadual do Oeste do Paraná, Brasil
Cláudia Regina Brescancini, Pontifícia Universidade Católica – RS, Brasil
Cristiane Lazzarotto-Volcão, Universidade Federal de Santa Catarina, Brasil
Cristine Gorski Severo, Universidade Federal de Santa Catarina, Brasil
Edair Maria Gorski, Universidade Federal de Santa Catarina, Brasil
Edwiges Maria Morato, Universidade Estadual de Campinas, Brasil
Fabio Luiz Lopes da Silva, Universidade Federal de Santa Catarina, Brasil
Felício Wessling Margotti, Universidade Federal de Santa Catarina, Brasil
Helena Guerra Vicente, Universidade de Brasília, Brasil
Heronides Maurílio de Melo Moura, Universidade Federal de Santa Catarina, Brasil
Izabel Christine Seara, Universidade Federal de Santa Catarina, Brasil
Josias Ricardo Hack, Universidade Federal de Santa Catarina, Brasil
Leandra Cristina de Oliveira, Universidade Federal de Santa Catarina, Brasil
Leonor Scliar Cabral, Universidade Federal de Santa Catarina, Brasil
Lucélio Dantas Aquino, Universidade Federal do Rio Grande do Norte, Brasil
Luizete Guimarães Barros, Universidade Federal de Santa Catarina, Brasil
Magdiel Medeiros Aragão Neto, Universidade Federal da Paraíba, Brasil
Mailce Borges Mota, Universidade Federal de Santa Catarina, Brasil
Maria Inêz Probst Lucena, Universidade Federal de Santa Catarina, Brasil
Maria Izabel de Bortoli Hentz, Universidade Federal de Santa Catarina, Brasil
Maria Teresa Santos Cunha, Universidade do Estado de Santa Catarina
Márluce Coan, Universidade Federal do Ceará, Brasil
Mary Elizabeth Cerutti-Rizzatti, Universidade Federal de Santa Catarina, Brasil
Maurício Eugênio Maliska, Universidade do Sul de Santa Catarina, Brasil
Monica Mano Trindade, Universidade Federal da Paraíba, Brasil
Morgana Fabiola Cambrussi, Universidade Federal da Fronteira Sul, Brasil
Nara Caetano Rodrigues, Universidade Federal de Santa Catarina, Brasil
Nelita Bortolotto, Universidade Federal de Santa Catarina, Brasil
Nívea Rohling, Universidade Tecnológica Federal do Paraná, Brasil
Otávio Goes de Andrade, Universidade Estadual de Londrina, Brasil
Renato Basso, Universidade Federal de São Carlos, Brasil
Roberta Pires de Oliveira, Universidade Federal de Santa Catarina, Brasil
Rodrigo Acosta Pereira, Universidade Federal de Santa Catarina, Brasil
Ronald Taveira da Cruz, Universidade Federal do Parnaíba Piauí, Brasil
Rosângela Hammes Rodrigues, Universidade Federal de Santa Catarina, Brasil
Rosely Xavier, Universidade Federal de Santa Catarina, Brasil
Simone Bueno Borges da Silva, Universidade Federal da Bahia, Brasil
Tarcisio de Arantes Leite, Universidade Federal de Santa Catarina, Brasil
Terezinha da Conceição Costa-Hübes, Universidade Estadual do Oeste do Paraná, Brasil
Vidomar Silva Filho, Instituto Federal de Santa Catarina, Brasil
Werner Heidermann, Universidade Federal de Santa Catarina, Brasil

Sumário

NÚMERO TEMÁTICO FALA, VOZ E EXPRESSIVIDADE

APRESENTAÇÃO	3
Letícia Rebollo Couto, Izabel Christine Seara	

ARTIGOS

UM ESTUDO ACÚSTICO DOS DITONGOS CENTRALIZANTES NA FALA CARIOCA	6
Pablo Arantes, Andrey Nikulin, Jessé da Silva Lima, Aveliny Mantovan Lima	
A VOZ COMO OBJETO DE UMA ANTROPOLOGIA DA ENUNCIÇÃO	35
Valdir do Nascimento Flores	
VOCAL ASPECT OF SOCIAL LAUGHTER DURING VIRTUAL INTERACTION	54
Takaaki Shochi, Marine Guerry, Hanako Suzuki, Mami Kanzaki, Jean-Luc Rouas, Toyoaki Nishida, Yoshimasa Ohmoto	
O PAPEL DA FONOLOGIA DO PORTUGUÊS BRASILEIRO NA TRADUÇÃO DE TEXTOS MUSICAIS	78
Graziela Pigatto Bohn, Roseane Silva da Cruz	
HOW PROSODY CONSTITUTES THE ACTIONS OF FORMULATIONS IN COURTROOMS	101
Ana Cristina Ostermann, Daniela Negraes Pinheiro Andrade, Minéia Frezza	
FRASEAMENTO PROSÓDICO DE ESTRUTURAS PARENTÉTICAS EM DADOS DE LEITURA NO PORTUGUÊS DO RIO DE JANEIRO	113
Vivian Borges Paixão, Carolina Ribeiro Serra	
MARCAÇÃO DE FOCO ESTREITO E O ACENTO SECUNDÁRIO EM INTERROGATIVAS TOTAIS NO PORTUGUÊS DO BRASIL	136
Manuella Carnaval, João Antonio de Moraes, Albert Olivier Blaise Rilliard	
"PRA CONQUISTAR A GAROTA...": DESGARRAMENTO E PROSÓDIA NAS VARIEDADES BRASILEIRA E LUSITANA DO PORTUGUÊS ..	168
Aline Ponciano dos Santos Silvestre	
A MULTI-DIRECIONALIDADE DA TRANSFERÊNCIA DA DURAÇÃO VOCÁLICA DO PORTUGUÊS COMO L3 PARA A L1 (ESPAÑHOL) E A L2 (INGLÊS): UM ESTUDO LONGITUDINAL	192
Letícia Pereyron, Ubiratã Kickhöfel Alves	
EFEITOS DE INSTRUÇÃO EXPLÍCITA NA AQUISIÇÃO DA ENTOAÇÃO EM ESPANHOL/L2	214
Rhanya Rafaella Rodrigues, Elena Ortiz Preuss	

IMAGEM DE BASE PARA A CAPA

TRÈS PARISIEN. LA MODE, LE CHIC, L'ÉLÉGANCE, ANONYMOUS, 1923

LETTERPRESS PRINTING, H 269MM × W 180MM

[HTTPS://WWW.RIJKSMUSEUM.NL/EN/COLLECTION/BI-1966-1051-8-2](https://www.rijksmuseum.nl/en/collection/BI-1966-1051-8-2)

APRESENTAÇÃO

Letícia Rebollo Couto | [Lattes](#) | leticiarcouto@yahoo.fr
Universidade Federal do Rio de Janeiro

Izabel Christine Seara | [Lattes](#) | izabel.seara@gmail.com | izabel.seara@ufsc.br
Universidade Federal de Santa Catarina

A fala e a voz humana não deixam de nos surpreender e de nos desafiar nas suas múltiplas acepções e possibilidades de estudo. Desde Jacques Cocteau com sua peça de teatro tantas vezes encenada e adaptada “La voix humaine” (1930) até Pierre Léon com seu tratado de fonostilística publicado na França em 1993, a voz e seus múltiplos aspectos de materialidade fônica e subjetiva têm sido importantes objetos de pesquisa e de criação estética, cruzando campos da interdisciplinariedade. O objetivo das editoras com este número foi dar um breve panorama das pesquisas realizadas sobre voz no cenário nacional e internacional, enquanto campo de conhecimento plural.

Trabalhos sobre os diferentes domínios das variações fônicas foram acolhidos pelas editoras que atuam em diversas áreas de interfaces, em particular no campo da prosódia, da interação conversacional, da variação dialetal e da expressividade, daí o tema **Fala, Voz e Expressividade**. A quantidade de artigos submetidos nos levou a considerar a edição de dois números, o primeiro de 2018-2 com o tema original da chamada e o segundo de 2019-1 com artigos mais específicos sobre as bases fonéticas necessárias para o estudo da fala e sobre as diversas **Interfaces da Prosódia**.

A origem das contribuições também é variada, incluindo jovens pesquisadores brasileiros, pesquisadores consolidados e líderes de grupos de pesquisa bem como contribuições internacionais, fruto de parcerias e colaborações que marcam a internacionalização das nossas equipes na área do domínio fônico e da expressividade. Os artigos deste volume contemplam temas e enfoques diversos, questões enunciativas, antropológicas, de fonologia do português, estrutura da informação e aquisição de línguas. O detalhe fonético está presente na maioria dos trabalhos, tanto nos seus aspectos acústicos quanto perceptuais, e nas suas consequências fonológicas para o que diz respeito à descrição do português do Brasil (PB). Ao todo dez artigos compõem este volume dedicado a repensar algumas questões sempre atuais como as de Ivan Fónagy (1991): quais são as especificidades da voz enquanto sistema de comunicação, sobre que bases se fundamenta a comunicação vocal indo além, bem além da mensagem que contém as palavras?

Para o campo da **Fala**, apresentamos o texto intitulado (1) *Um estudo acústico sobre ditongos centralizantes na fala carioca*, de autoria de Pablo Arantes, Andrey Nikulin, Jessé da Silva Lima e Aveliny Mantovan Lima. Ditongos centralizantes são definidos como vogais que, apesar de serem, em nível subjacente, monotongos, são percebidas como uma sequência de vogal mais uma semivogal de qualidade central.

Para o campo da **Voz**, foram considerados estudos aplicados à antropologia da enunciação, como o texto intitulado (2) *A voz como objeto de uma antropologia da enunciação*, de autoria de Valdir do Nascimento Flores, que apresenta uma análise de comentários acerca da voz de cantores brasileiros. Ainda sob o escopo da Voz, temos o texto intitulado (3) *Vocal aspect of social laughter during virtual interaction*, de autoria de Takaaki Shochi, Marine Guerry, Hanako Suzuki, Mami Kanzaki, Jean-Luc Rouas, Toyoaki Nishida e Yoshimasa Ohmoto, que trata da descrição acústica e do reconhecimento do riso espontâneo e social, gravado em situações de imersão digital e reconhecido por ouvintes franceses e japoneses nas modalidades áudio (A) e audiovisual (AV). Este estudo demonstra que, do ponto de vista acústico, o riso espontâneo diferencia-se do riso social pela duração e pela duração do vozeamento. Ambos são bem reconhecidos por ouvintes japoneses e franceses. Entretanto, do ponto de vista da avaliação de fatores como estado físico, envolvimento do falante e distância psicológica, há divergência entre a avaliação de franceses e japoneses quanto à distância psicológica. Na sequência, apresentamos o texto intitulado (4) *O papel da Fonologia do português brasileiro na tradução de textos musicais*, de Graziela Pigatto Bohn e Roseane Silva da Cruz. Com base no ritmo, acento e nos processos de junção segmental, sob o enfoque fonológico, e nos conceitos de acento métrico e ritmo, sob o enfoque da música, as autoras pesquisam “o processo tradutológico de músicas para o canto”.

Para o campo da **Expressividade**, foram considerados estudos prosódicos com interface sintática ou pragmática relacionados aos atos de fala, atitudes e/ou emoções. Nesse campo, o primeiro texto intitulado (5) *How prosody constitutes the actions of formulations in courtrooms*, de autoria de Ana Cristina Ostermann, Daniela Negraes Pinheiro Andrade e Minéia Frezza, apresenta a prosódia contribuindo para dar sentido às ações realizadas em audiências de instrução. Na sequência, temos três textos relacionados à descrição do português e dois textos relacionados a temas de aquisição de línguas estrangeiras ou adicionais. O texto intitulado (6) *Fraseamento prosódico de estruturas parentéticas no português do Brasil*, de Vivian Borges Paixão e Carolina Ribeiro Serra, que apresenta, a partir do viés fonológico, uma investigação sobre as estruturas parentéticas, contribuindo

para a descrição da gramática entoacional do PB, em relação ao fraseamento prosódico. O texto (7) *Marcação de foco estreito e o acento secundário em interrogativas totais no português do Brasil*, de Manuella Carnaval, João Antonio de Moraes e Albert Rilliard, descreve as realizações de foco em *corpus* de fala atuada, considerando fatores de variação como o número de sílabas e posição. A descrição contempla aspectos acústicos e perceptuais do foco e do acento secundário. O texto (8) *“Pra conquistar a garota...”: Desgarramento e Prosódia nas variedades brasileira e lusitana do Português*, de Aline Ponciano dos Santos Silvestre, propõe uma análise fonológica de orações adverbiais desgarradas, comparando duas variedades do português. A marcação fonológica proposta difere no português brasileiro e no português europeu, bem como a implementação da duração.

A duração é também um elemento distintivo no trabalho apresentado no texto (9) *A multi-direcionalidade da transferência da duração vocálica do português como L3 para a L1 (espanhol) e a L2 (inglês): um estudo longitudinal*, de Letícia Pereyron e Ubiratã Kickhöfel Alves. Esse trabalho tem como foco a comparação da implementação da duração em português (L3), inglês (L2) e espanhol (L1) nas vogais tônicas antes e após instrução fornecida ao longo de quatro meses. O estudo demonstra que as alterações nas vogais do português (L3) causam alterações nas durações vocálicas das demais línguas devido à interconexão dos sistemas do falante multilíngue. E finalmente no texto (10) *Efeitos de Instrução Explícita na Aquisição da Entoação em Espanhol/L2*, de Rhanya Rafaella Rodrigues e Elena Ortiz Preuss, são discutidos aspectos prosódicos da aquisição da entoação em Espanhol/L2. As autoras concluem que a dificuldade na produção de enunciados interrogativos totais é maior do que a de enunciados declarativos em termos de aquisição.

Pela novidade de algumas das interfaces apresentadas neste número, bem como pelo desdobramento de trabalhos no próximo número **Interfaces da Prosódia**, esperamos consolidar as relações e interrelações entre a linguística e outros campos de conhecimento aplicados a diversas questões relacionadas ao lugar da voz e da fala na sociedade.

Referências

- FÓNAGY, Ivan. *La vive voix, essai de psycho-phonétique*. Préface de Roman Jakobson, Payot, 1991.
- LÉON, Pierre Roger, *Précis de phonostylistique : parole et expressivité*. Paris : Nathan Université, 1993. [Série « linguistique »].

UM ESTUDO ACÚSTICO DOS DITONGOS CENTRALIZANTES NA FALA CARIOCA

AN ACOUSTIC STUDY OF THE CENTERING DIPHTHONGS
IN THE RIO DE JANEIRO VARIETY OF BRAZILIAN PORTUGUESE

Pablo Arantes | [Lattes](#) | pabloarantes@gmail.com
Universidade Federal de São Carlos

Andrey Nikulin | [Lattes](#) | andre.n.guzman@gmail.com
Universidade de Brasília

Jessé da Silva Lima | [Lattes](#) | jesse.1548@gmail.com
Universidade de Brasília | CNPq

Aveliny Mantovan Lima | [Lattes](#) | avelinylima@gmail.com
Universidade de Brasília

Resumo: O objetivo do trabalho é fazer a caracterização acústica de ocorrências daquilo que chamamos de ditongos centralizantes (DC) na fala carioca. Definimos DC como vogais que, embora subjacentemente sejam monotongos, são percebidas como uma sequência de vogal seguida de uma semivogal com qualidade central. Ocorrências de DC e de monotongos realizados como tal foram buscadas em um *corpus* de entrevistas radiofônicas. As trajetórias formânticas e a duração dos dois tipos de sequências foram comparadas visualmente e quantitativamente por meio da transformada discreta do cosseno. Os resultados indicam que há diferenças marcantes entre os dois tipos de vogais. A trajetória dos DC apresenta movimentos de maior magnitude nas dimensões horizontal e vertical do espaço vocálico, que se iniciam na região mais periférica e terminam na região central, em alturas variadas. Os DC são também mais longos do que os monotongos, quer estejam em posição prosódica proeminente ou não. Os dados de contagem de ocorrências de DC sugerem que há uma preferência pela ocorrência desses segmentos em posição de proeminência prosódica e em momentos de disfluência. Apontamos que há a possibilidade de este viés distribucional evoluir para o estabelecimento de um fenômeno de distribuição complementar entre monotongos e DC. Sugerimos, por fim, com base na Fonologia Articulatória, que a emergência da articulação centralizante resulta da amplificação, em contextos que induzem o alongamento vocálico, da sincronia imperfeita entre os gestos de corpo de língua e glotal.

Palavras-chave: Português brasileiro; Fala carioca; Ditongo centralizante; Fonética acústica.

Abstract: In this study, we offer an acoustic characterization of the occurrences of so-called centering diphthongs (CD) in the Rio de Janeiro variety of Brazilian Portuguese. We define CD as underlying monophthongs that surface as sequences of a vowel followed by a central semivowel (inglide). CD and plain monophthongs were manually segmented from a radio interview corpus, and their formant trajectories and duration were subsequently compared both visually and quantitatively, using the technique of the discrete cosine transform. Our results show that there are striking differences between these two types of vowels. The articulatory gestures involved in CD production are of greater magnitude in horizontal and vertical dimensions of the vowel space; they originate in a peripheral region and are directed towards the central region at varying heights. The duration of the CD is also longer than that of plain monophthongs, regardless of whether the environment is prosodically prominent. A distributional analysis shows that there is a preference for the occurrence of CD in prosodically prominent environments, as well as in the context of disfluencies. We suggest that this distributional bias may contribute to the emergence of a complementary distribution between plain monophthongs and CD. Finally, we suggest, based on Articulatory Phonology, that centering glides in PB can emerge as a result of imperfect synchrony between the tongue body and glottal gestures that gets enhanced in lengthening-inducing environment.

Keywords: Brazilian Portuguese; Rio de Janeiro dialect; Centering diphthongs; Acoustic phonetics.

1. Introdução

Embora facilmente reconhecível pelos falantes do português brasileiro (PB), a variedade da língua utilizada na região metropolitana do Rio de Janeiro (daqui em diante *português carioca*) carece de estudos acústicos de algumas de suas características marcantes. Neste trabalho oferecemos uma descrição fonética preliminar do que vamos chamar de “ditongos centralizantes” (DC), fenômeno que consiste na realização ditongada opcional de algumas vogais, especialmente em posição tônica, que poderiam, preliminarmente, ser transcritas como segue: *amiga* [a' miɣg̊], *porta* ['pɔɣxtɐ], *até* [a' tɛɣ̊]¹. Não se enquadram nessa descrição, e, portanto, não serão tratados aqui, casos bastante comuns nessa e em outras variedades do PB de assimilação regressiva, como a iotização, em palavras como *vez* ['vej̯]), nos quais a vogal assimila o traço palatal da consoante seguinte.

¹ Ressaltamos que a escolha do [ɣ̊] como símbolo para a transcrição da realização centralizada dessa ditongação reflete apenas uma análise preliminar de oitiva e não há, nesse momento, um compromisso de correspondência com a exatidão dessa vogal no espaço vocálico.

No presente trabalho, com o objetivo de dar uma caracterização fonético-acústica para os DC, analisamos a trajetória das vogais /i, e, ε, a, ə, o, u/ de um *corpus* de fala espontânea, considerando tanto as realizações prototípicas (monotongos tônicos [i, e, ε, a, ə, o, u] e pós tônicos [ɪ, ɐ, ʊ]), como as realizações que apresentaram o fenômeno de DC. Não foram encontrados exemplos de DC para todas as vogais.

1.1. O sistema vocálico do PB

É consenso que o sistema vocálico do PB compreende 7 fonemas (/a, ə, o, u, ε, e, i²) (CÂMARA JR., 1970, p. 33; SEARA et al., 2011, p. 45), sendo que o *status* fonológico é geralmente negado às vogais nasais (CÂMARA JR., 1970, p. 49; CAGLIARI, 1977; LÓPEZ, 1979; WETZELS, 1991, 1997; MORAES; WETZELS, 1992). A estes fonemas correspondem os seguintes fones, encontrados na posição tônica: [a, ə, o, u, ε, e, ɪ, ẽ, ã, õ, ù, ẽ, ĩ].³

O PB permite a ocorrência de múltiplos segmentos vocálicos adjacentes, formando encontros vocálicos, que podem ser ora heterossilábicos (por exemplo, *caô* [ka.'o], *Caim* [ka.'ĩ]), ora tautossilábicos (por exemplo, *mau/mal* [maʊ], *muito* ['mũĩ.tu]). Embora os encontros vocálicos tautossilábicos sejam frequentemente chamados de *ditongos*, não adotaremos este uso no presente trabalho, reservando o termo “ditongo” somente às sequências vocálicas tautossilábicas que não permitem uma análise bissegmental. Como o fenômeno de DC cria segmentos de contorno passíveis de uma análise monossegmental, neste trabalho optamos por usar o termo “ditongo” para nos referirmos a eles.

1.2. Estudos dos processos de ditongação e monotongação no PB

Muitas variedades regionais e socioletais do PB apresentam fenômenos de ditongação de monotongos ou monotongação de encontros vocálicos. Entre os casos mais bem estudados, principalmente dentro do paradigma variacionista, destacam-se:

- a realização opcional de [ẽ, ã] como [e, a], principalmente diante de [ɾ, ʃ, ʒ], e a realização opcional de [õ] como [o] em todos os ambientes (RECTOR, 1975; VEADO, 1983; MOTA, 1988; BISOL, 1994; CABREIRA, 1996; PAIVA, 1996, 2003; MOLLICA, 1998; ARAÚJO, 1999; LOPES, 2002; SILVA, 2004; PEREIRA, 2004; DUARTE; PAIVA, 2011; SEARA et al., 2011, p. 43-44; LEÃO, 2013; ARAGÃO, 2014);

² Não discutiremos aqui a possibilidade de atribuir o *status* fonêmico ao segmento [ɜ], encontrado em empréstimos recentes e nomes próprios como *hora do rush*, *Ruffles* ou *Pizza Hut* na fala de muitos falantes do PB.

³ [ẽ] é transcrito como [ɜ̃] por alguns autores.

- a inserção do *offglide* [ɪ] antes de /s/ em coda ou diante de fricativas pós-alveolares heterossilábicas (ALBANO, 2001; LEITE et al., 2003; COLLEY, 2009; ALMEIDA, 2009, p. 85, 88; ARAGÃO, 2014; SILVA, 2014; ROCHA et al., 2015);
- a realização de [õ] como [õõ] (ALMEIDA, 2009, p. 96);
- a realização de [ẽ] como [ẽĩ] (OUSHIRO; MENDES, 2014).

No entanto, os processos que envolvem a formação de ditongos decrescentes em que o movimento articulatório do complexo línguo-mandibular segue para uma posição mais baixa não foram suficientemente bem estudados. Além de um estudo sobre a realização opcional de [ẽ] como [ẽõ] por falantes da região metropolitana do Rio de Janeiro (ver COLLEY, 2011), encontramos apenas os estudos referentes à realização ditongada das vogais tônicas orais na variedade do PB falada em Porto Alegre (RS) (ver BATTISTI, 2013; BATTISTI; OLIVEIRA, 2014, 2016).

1.3. Objetivo e estrutura do trabalho

O objetivo buscado será descrever a realização ditongada de vogais da variante carioca do português brasileiro. Para isso, iremos contrapor vogais que se apresentam como DC e outras que são realizadas como monotongos convencionais.

O restante deste trabalho está estruturado da seguinte maneira: na seção 2, discutiremos a tipologia de processos de ditongação em diversas línguas do mundo com um enfoque na criação de ditongos centralizantes; a seção 3 abordará a metodologia utilizada neste trabalho, explicitando os materiais linguísticos, as variáveis definidas e o aparato instrumental empregado para efetuar as análises acústicas e estatísticas; na seção 4, discutiremos os resultados e o panorama que eles indicarem a partir dos dados obtidos; por fim, a seção 5 concluirá o estudo e apontará possibilidades de pesquisas futuras.

2. Ditongação diacrônica: o movimento em direção ao centro

Para um melhor entendimento do processo a ser analisado neste trabalho, é importante levar em consideração que um incremento de duração de uma vogal pode favorecer a criação de segmentos de contorno. De fato, Battisti (2013) já observou que o fenômeno de *ingliding* na fala porto-alegrense é registrado nas vogais com realização alongada. No entanto, Battisti e Oliveira (2014, 2016) argumentam que o *ingliding* pode ocorrer em vo-

gais tanto alongadas quanto não alongadas, sendo que o alongamento facilita a percepção de oitiva do *ingliding*. Segundo os autores, na variedade porto-alegrense do PB o *ingliding* nas vogais não alongadas pode ser detectado apenas via análise acústica.

Contudo, as evidências diacrônicas de outras línguas mostram que os ditongos decrescentes, fonêmicos ou não, frequentemente emergem como um resultado de um processo diacrônico de ditongação (*vowel breaking*) de monotongos alongados (tipicamente se trata de vogais médias). Esse desenvolvimento, quando não causado por processos de *umlaut*, tende a seguir um dos dois caminhos possíveis.

A primeira possibilidade envolve a retenção, ainda que parcial, da região articulatória e do arredondamento dos segmentos originais, conforme ilustrado na Tabela 1 abaixo.

Tabela 1. Criação de ditongos decrescentes não-centralizantes a partir de monotongos

língua	alvo	fonte	referência
finlandês	iɛ̯, uɔ̯, yø̯ ⁴	*e:, *o:, *ø:	ITKONEN, 1983, p. 374; HAKULINEN et al., 2004, p. 55-56
iacute	iɛ̯, uɔ̯, yø̯ ⁵	*e:, *o:, *ø:	TENIŠEV; DYBO, 2006, p. 224-225
latgalês	iɛ̯, uɔ̯	*e:, *o:	NAU, 2011, p. 9; SCHMALSTIEG, 1961
elfdaliano	iɛ̯, uɔ̯, yœ̯ ⁶	*eɪ, *o:, *ø:	SAPIR, 2005
iucaguir da tundra, iucaguir de Kolymá	iɛ̯, uɔ̯, yø̯ ⁷	*e:, *o:, *ø:	MUDRAK, 2010, p. 167

Fonte: Elaborada pelos autores.

Na história de outras línguas, as vogais médias (geralmente longas) se transformaram em ditongos centralizantes. Nos ditongos desse tipo observa-se um movimento articulatorio para a região central, geralmente acompanhado pela perda do arredondamento, conforme ilustrado na Tabela 2 abaixo.

⁴ Em algumas variedades regionais do finlandês o movimento articulatorio na produção destes três ditongos chegou a fazer um percurso ainda maior, atingindo o grau de abertura mais baixo: [iɛ̯ ~ iæ̯ / iɔ̯, uɔ̯ ~ uæ̯, yæ̯] (HAKULINEN et al., 2004, p. 55-56). Contudo, trata-se de um desenvolvimento recente, cujo ponto de partida não foram as vogais médias longas e sim os ditongos decrescentes com um offglide médio.

⁵ Este processo foi acompanhado pela formação do ditongo [iɔ̯] a partir de sequências heterossilábicas.

⁶ Segundo Sapir (2005, p. 20), /uo/ pode ser realizado como [yæ̯] em algumas variedades. Os três ditongos podem ainda ser nasalizados ou alongados.

⁷ Nessas duas línguas existe ainda um ditongo [iɔ̯] de origem incerta.

Tabela 2. Criação de ditongos decrescentes centralizantes a partir de monotongos

língua	alvo	fonte	referência
lituano	i̯, u̯	*e:, *o:	BALODE; HOLVOET, 2001b, p. 46; SCHMALSTIEG, 1961
frísio ocidental	i̯, i̯, u̯, o̯ ⁸	*e:, *ɛ:, *o:, *ɔ:	TIERSMA, 1983, p. 60
luxemburguês	i̯, u̯	*e/*ɛ, *a/*o	GILLES; TROUVAIN, 2013

Fonte: Elaborada pelos autores.

Há ainda uma terceira possibilidade, na qual as duas estratégias de ditongação co-atuam. Nas poucas línguas que apresentam esse tipo de desenvolvimento diacrônico, algumas vogais desenvolvem um *offglide* na mesma região articulatória, enquanto outras criam um *offglide* centralizante. Isso está exemplificado na Tabela 3.

Tabela 3. Criação de ditongos centralizantes e não-centralizantes a partir de monotongos diferentes

língua	alvo	fonte	referência
letão	i̯, u̯	*e:, *o:	BALODE; HOLVOET, 2001a, 11; SCHMALSTIEG, 1961
africâner	i̯, u̯, i̯ ⁹	*e:, *o:, *ø:	LASS, 1987

Fonte: Elaborada pelos autores.

Neste trabalho, procuraremos situar o português carioca nesta classificação, demonstrando que o processo de ditongação nessa variedade do PB cria ditongos centralizantes (como no lituano, no frísio ocidental e no luxemburguês), possivelmente a partir dos alofones alongados por fatores prosódicos ainda não estudados. O processo, ainda em andamento e, portanto, ainda não incorporado aos contrastes fonêmicos da variedade em questão, teria atingido as vogais médias-baixas, médias-altas e altas, mas não a vogal baixa /a/.

⁸ O frísio possui, ainda, os ditongos [y̯, ø̯], de baixa ocorrência. Estes provavelmente provêm de *ɔ:, *œ:, embora este fato não seja mencionado na fonte consultada.

⁹ Lass (1987) registra um alto grau de variação fonética para esses fonemas: [i̯ ~ i̯ ~ i̯], [u̯ ~ u̯ ~ u̯], [ø̯ ~ ø̯].

3. Metodologia

Nesta seção apresentaremos o material linguístico utilizado na presente pesquisa, incluindo os critérios de recorte adotados (3.1), assim como o aparato tecnológico e estatístico (3.2) empregado para a medição, quantificação e análise dos dados selecionados.

3.1. Material de fala

O *corpus* utilizado neste trabalho foi constituído por gravações de fala espontânea, em sua maioria colhidas de *podcasts* gravados pela rádio CBN (disponíveis no site www.cbn.globoradio.globo.com). Essa prática foi adotada de forma bem-sucedida em trabalhos como Kohler (1995) e Ostendorf et al. (2001), tendo a vantagem de permitir a recolha rápida e barata de exemplos de fala bastante diversos em gravações de duração relativamente curta.

Os áudios recolhidos geralmente têm como temática a profissão do sujeito convidado, como, por exemplo, a carreira jornalística, a prática médica ou a produção de energia renovável, além das vivências pessoais de cada indivíduo. Os trechos em que o entrevistador participa da conversa foram descartados para que os possíveis achados não pudessem ser atribuídos ao fonoestilo jornalístico.

As amostras de fala do *corpus* provêm de um total de 13 falantes, sendo oito mulheres e cinco homens, todos naturais da região metropolitana do Rio de Janeiro, de condição social elevada. Em termos de escolaridade, todos apresentam ensino médio ou superior completo. A duração combinada de todas as amostras do *corpus* totaliza 15m55.

3.1.1. Critérios para a seleção dos dados

Como critério para decidir quais dados de fato seriam analisados neste trabalho, optamos por excluir variáveis que poderiam interferir no fenômeno de DC. A fim de evitar a necessidade de uma análise multifatorial, consideramos apenas as vogais que em seu contexto posterior não apresentavam consoantes coronais (ou seja, cuja articulação não envolvia ponta de língua). Com esta delimitação, procuramos minimizar a potencial interação com um fenômeno de ditongação distinto, descrito por autores como Albano (2001, p. 86-91), Leite et al. (2003), Colley (2009), Almeida (2009, p. 85, 88), Aragão (2014), Silva (2014), Rocha et al. (2015), entre outros. Albano (2001) analisa as realizações ditongadas de palavras com o arquifonema /S/ em coda final, como por exemplo “três” e “arroz”, que ela denomina *iotização*, e postula a inserção de [j] diante de /s/.

Segundo a autora, embasada na fonologia articulatória de Browman e Goldstein (1989), a iotização seria motivada não apenas por questões prosódicas, mas também por razões articulatórias:

A lentificação da transição implica uma antecipação da trajetória variável do trato local de constrição da ponta de língua (LCPL) em direção ao alvo. Ao mover-se lentamente, a ponta da língua leva consigo o corpo da língua como articulador passivo. Esse, então, desencadeia a iotização ao passar pelo palato duro numa velocidade apenas um pouco superior à de uma semivogal. (ALBANO, 2001, p. 89)

Dessa forma, escolhemos analisar as vogais seguidas por consoantes bilabiais, labiodentais, velares (incluindo as realizações velares¹⁰ dos róticos do PB), além de considerar também aquelas que precederam pausas. Ademais, decidimos evitar contextos fônicos que pudessem atrapalhar a medição acústica dos formantes, como a nasalidade, em razão das dificuldades relacionadas à identificação automatizada de F_1 e F_2 em trechos afetados por essa característica acústico-articulatória (STEVENS, 1998, p. 303-322).

3.1.2. Rotulagem dos dados

Para um melhor manejo do *corpus*, as vogais e os encontros vocálicos selecionados para a análise foram etiquetados, sendo que os rótulos incluem, obrigatoriamente, a informação sobre as seguintes variáveis:

- sujeito de fala (número de identificação);
- sexo do sujeito de fala (m = feminino, h = masculino);
- tipo de sequência segmental (mt = monotongo, ca = ditongo carioca, dt = encontro vocálico tautossilábico);
- ambiente posterior (ps = pausa, R = rótico velar/pós-velar em coda, rr = rótico velar/pós-velar em ataque, bb = bilabial, lb = labiodental);
- presença da proeminência frasal (s = presente, n = ausente, sh = hesitação);¹¹
- representação das vogais segmentadas (a = /a/ tônica, A = /a/ pós-tônica, O = /ɔ/ tônica, o = /o/ tônica, u = /u/ tônica, U = /u/ pós-tônica, E = /ɛ/ tônica, e = /e/ tônica, i = /i/ tônica, I = /i/ pós-tônica).

¹⁰ De acordo com a análise de oitava de nossos dados, no português carioca o rótico em coda é realizado predominantemente com uma fricção característica de fricativas pós-velares e não velares (*contra* CALLOU et al., 1998). No entanto, uma análise mais detalhada foge do escopo do presente trabalho.

¹¹ Optamos por distinguir as instâncias de proeminência que ocorrem em trechos fluentes (rótulo s) das ocorrências de hesitação e das marcas discursivas (rótulo sh), associadas com a fala disfluente.

É necessário ressaltar que, dada a inexistência de um critério acústico objetivo para a identificação de um DC (uma aproximação desta definição é um dos objetivos deste trabalho), a classificação dos vocoides selecionados para análise em monotongos ou DC foi realizada de oitiva. Dois dos autores do trabalho fizeram a classificação de forma independente. Nos casos em que houve discordância, os outros dois autores foram consultados a fim de chegar-se a um consenso por meio da audição conjunta. Uma vez que havia apenas dois juízes fazendo a classificação independente e os casos de discordância não passaram de 10%, julgamos que não era o caso de usar um teste estatístico como kappa para quantificar as discrepâncias. O mesmo procedimento foi aplicado em relação à proeminência frasal. O critério estabelecido para classificar um vocoide como DC foi a identificação de uma realização na qual houvesse mudança de qualidade perceptível ao longo de sua articulação. Quanto à proeminência frasal, foram rotuladas como proeminentes (rótulo *s*) as vogais tônicas das palavras que se destacavam dentro de um trecho entre pausas. As vogais que ocorriam em trechos disfluentes (por exemplo, aquelas que faziam parte de marcadores discursivos ou de hesitações) foram rotuladas à parte (rótulo *sh*).

Na Tabela 4, apresentamos o número de ocorrências de cada vogal nos dados selecionados para a análise.

Tabela 4. Número de instâncias analisadas de cada vogal

vogal	posição	DC	monotongo
/i/	tônica	11	22
	pós-tônica	1	2
/e/	tônica	23	6
/ɛ/	tônica	8	17
/u/	tônica	1	6
	pós-tônica	0	4
/o/	tônica	2	11
/ɔ/	tônica	17	7
/a/	tônica	0	21
	pós-tônica	0	1

Fonte: Elaborada pelos autores.

A ausência total das ocorrências de DC na vogal /a/ já foi atestada por Battisti e Oliveira (2014, p. 47; 2016, p. 16). Uma possível explicação desta lacuna distribucional

é que, conforme mostraremos na seção 4, o grau de abertura dos DC aumenta dinamicamente no *offglide*, o que seria impossibilitado no caso da vogal /a/, já que a realização padrão dessa vogal no PB envolve a maior abertura dentre as vogais da língua.

3.2. Análise acústica

Nesta seção detalharemos os procedimentos adotados para a análise acústica dos dados, discutindo a análise da duração (3.2.1), dos valores de formantes (3.2.2), os procedimentos de normalização (3.2.3) e a análise estatística (3.2.4).

3.2.1. Análise de duração

A duração acústica dos segmentos foi determinada a partir dos instantes de início e fim dos segmentos, que foram identificados manualmente e marcados em arquivos de metadados do *Praat* (BOERSMA, WEENINK, 2017) para posterior extração automatizada por meio de um *script* do mesmo programa. A combinação de informações visuais presentes no oscilograma e no espectrograma do sinal acústico foram usados para identificar os instantes de início e fim dos vocóides. O instante de início está associado ao aumento súbito na amplitude do oscilograma e no aparecimento de energia nas frequências altas do espectrograma, indicando a presença clara dos formantes superiores (F_2 e maiores). O instante de fim foi associado à diminuição rápida da amplitude do oscilograma e no esmaecimento da presença dos formantes superiores no espectrograma. A duração bruta dos segmentos passou pelo procedimento de normalização por *z-score* para tornar possível a comparação conjunta de todas as qualidades vocálicas sem a interferência da duração intrínseca. A técnica usada para fazer a normalização da duração é o *z-score* estendido, assim como apresentada detalhadamente em Barbosa (2006). Para a aplicação desse procedimento, recorre-se a uma tabela, apresentada no apêndice B.1 de Barbosa (2006), com valores de média e desvio-padrão para os fones do PB que são usados como referência. Os valores foram calculados a partir de um *corpus* de fala lida, no qual cada segmento do PB apareceu em todos os contextos fônicos admitidos pela fonotaxe da língua. A normalização da duração é feita segundo a fórmula

$$z = \frac{dur - \mu}{\sigma},$$

onde z é o valor normalizado, dur é a duração a ser normalizada, expressa em unidade de tempo, de um segmento e μ e σ são, respectivamente, os valores da média e do desvio-padrão extraídos da tabela de referência para uma vogal do mesmo tipo daquela a ser normalizada.¹² Assim, se o dado a ser normalizado é a duração de uma vogal [i], os valores μ e σ usados na fórmula serão os de uma vogal [i]. No caso das vogais classificadas perceptualmente como DC, os valores de referência usados para fazer a normalização foram os do monotongo correspondente. Esse procedimento foi adotado pelas seguintes razões: (a) os DC não têm, em PB, valor fonológico distinto dos monotongos; (b) comparar os casos classificados como DC com a referência dos monotongos permitirá saber se os DC são caracterizados por um alongamento sistemático relativamente às vogais que não receberam essa classificação; (c) a tabela de referência disponível para os sons do PB lista os ditongos com valor fonológico, que são sempre *offglides*, terminados nas semivogais [i] e [u] e não há razão a priori nem fonética nem fonológica para supor que se possa pôr em equivalência os chamados ditongos verdadeiros do PB com as vogais classificadas aqui como DC. O procedimento de normalização da duração bruta foi automatizado por meio de um *script* para o ambiente de computação estatística R.

A mesma etiquetagem discutida na seção 3.1.2 foi aproveitada para a análise de duração dos segmentos classificados como monotongos ou DC. O interesse no parâmetro de duração está na comparação entre as durações de monotongos e DC e entre os segmentos marcados como proeminentes ou não proeminentes frasalmente, além da interação entre essas duas variáveis.

¹² Um dos pareceristas anônimos sugere que um valor de z -score muito alto poderia indicar que a ocorrência do alongamento indicado por esse valor poderia ser considerada improvável caso se interpretasse o valor de z como uma probabilidade. Embora a transformação de dados brutos em z -scores tenha origem na estatística, não nos parece que Barbosa (2006) sugira que os valores de duração normalizados através dessa transformação devam ser interpretados como probabilidades, mesmo porque, para isso, média e desvio-padrão de pausa silenciosa deveriam entrar para o cálculo de um z -score estatístico. Esses valores dados pela normalização são um índice descritivo de alongamento ou encurtamento de uma determinada unidade silábica relativamente a descritores de tendência central e dispersão típicos dos fones que a compõem. Exemplificando para o caso de fones, consideramos que um fone com valor z de 1,5 sofreu um alongamento menor do que outro com valor z de 3, mas não derivamos daí que o primeiro alongamento seja estatisticamente mais provável do que o segundo. Interpretar os valores de z -score como probabilidades no contexto da normalização de duração assim como proposta por Barbosa (2006) introduziria também a questão da distribuição de probabilidade a ser usada como referência para julgar um valor empírico de z . O uso de valores z para fazer inferências supõe uma distribuição normal, suposição que não é necessariamente válida. Barbosa (2006) e Rosen (2005) mostram que durações segmentais de vogais e consoantes são melhor modeladas por distribuições não-normais, como lognormal e gama.

3.2.2. Extração dos valores de formantes

Para os formantes, o intervalo de análise foi o mesmo definido para a duração. O *script* FormantPro (XU, GAO, 2018) foi usado para extrair a trajetória normalizada temporalmente dos valores de F_1 , F_2 e F_3 das vogais do *corpus*. Especificamos a coleta de 20 amostras igualmente separadas no tempo a serem extraídas em cada trecho. Para as falantes femininas, utilizamos os valores padrão sugeridos para os parâmetros da técnica LPC, usada pelo *script* para a extração dos formantes com frequência máxima de 5,5 kHz e um máximo de 5 formantes. Para os falantes masculinos, fixamos a frequência máxima em 5 kHz e um máximo de 5 formantes. Verificamos posteriormente que, para algumas das ocorrências, especialmente de /i e u/, essas especificações produziram extrações errôneas. Nesses casos, aplicamos o procedimento de variar separadamente os valores dos dois parâmetros relevantes de extração até que fosse possível conseguir uma extração aceitável. Nenhum dado foi descartado por ter sido impossível conseguir uma extração adequada.

3.2.3. Procedimento de normalização

Como o *corpus* é formado por amostras de fala de múltiplos falantes, tanto do sexo feminino quanto masculino, foi necessário recorrer a uma técnica de normalização dos valores dos formantes para que fosse possível analisar conjuntamente todos os dados. Em função das características do *corpus*, não foi possível obter para todos os falantes exemplos de todas as vogais do inventário da língua. Por conta disso não foi possível aplicar as técnicas de normalização do tipo *vowel-extrinsic* mais conhecidas, como as de Lobanov (1971) e Nearey (1978), por exemplo. Por conta dessa limitação, recorreremos a uma técnica de tipo *vowel-intrinsic*, que pode ser aplicada a partir de informações obtidas para cada ocorrência de vogal no *corpus*. Recorreremos a um procedimento proposto por Syrdal e Gopal (1986), que consiste em representar a dimensão do avanço da língua no eixo anterior-posterior e a dimensão da altura por meio de diferenças entre valores de formantes convertidos da escala física Hertz para a escala perceptual Bark. Seguindo sugestão de Thomas (2011, p. 163), que modifica a proposta original de Syrdal e Gopal, a dimensão do avanço é dada pela diferença $Z_3 - Z_2$ e a da altura por $Z_3 - Z_1$, em que Z_i é o valor do i -ésimo formante convertido da escala Hz para a escala Bark. Para a conversão, empregamos a fórmula

$$Z_i = \frac{26,81}{1 + \frac{1960}{F_i} - 0,53},$$

apresentada em Thomas (2011, p. 58). O ambiente de computação estatística R (R CORE TEAM, 2017) foi usado para a conversão dos valores e demais operações necessárias para a análise das trajetórias formânticas.

A extração de 20 amostras ao longo de cada trecho analisado permite que avaliemos a evolução da trajetória formântica no espaço articulatório ao longo do tempo. No caso dos monotongos, esperamos que a trajetória seja basicamente estacionária, com pouco grau de movimentação. Nos casos identificados na análise de oitiva como ocorrências de ditongos centralizantes, esperamos encontrar evidência de um grau comparativamente maior de movimentação. A hipótese que levantamos é que, nesses casos, a trajetória tem início na região articulatória próxima àquela ocupada pelo monotongo de mesma qualidade vocálica e gradualmente se move em direção à região central do espaço vocálico.

A fim de testar essa hipótese, geramos a trajetória formântica média para os monotongos e para os casos de DC separadamente (vide Tabela 4 para as quantidades de ocorrências de cada qualidade vocálica no nosso *corpus*). Para evitar os efeitos mais pronunciados da coarticulação com os segmentos circundantes, desprezamos os dois primeiros e os dois últimos valores da série de vinte gerados pela aplicação do *script* FormantPro.

3.2.4. Análise quantitativa das trajetórias

A observação do traçado das trajetórias dos monotongos e DC no espaço vocálico, apresentados na seção 4, permite uma análise visual e qualitativa das possíveis diferenças entre ambos os tipos. Para permitir uma quantificação das diferenças entre os dois tipos que complementasse a análise qualitativa, recorreremos à técnica da transformada discreta do cosseno (*discrete cosine transform*, DCT, em inglês), da maneira sugerida por Watson e Harrington (1999). Conforme explicam os autores, a técnica permite modelar a trajetória temporal de um formante reduzindo-a a coeficientes da DCT.¹³ Esses coeficientes têm valores que são proporcionais à presença de diferentes componentes derivados de uma função-base, nesse caso a função cosseno. O primeiro coeficiente representa uma linha reta e seu valor é proporcional à média dos valores da trajetória. O segundo coeficiente quantifica a presença de um componente que é um meio ciclo da função cosseno. No presente contexto, ele pode ser entendido com uma medida da direção e da magnitude da

¹³ Um parecerista observa que existem outras técnicas que poderiam ser utilizadas com a mesma finalidade (isto é, a decomposição da trajetória formântica em componentes elementares), tais como os polinômios ortogonais (LEVITI; NEUMAN, 1991; GRABE et al., 2007) ou a análise funcional de dados (ASTON et al., 2010), entre outras. A escolha da DCT para o presente estudo não foi precedida por uma comparação sistemática de todas as técnicas disponíveis, mas pelo fato dela ser apropriada para os propósitos do trabalho.

inclinação da trajetória do formante. Assim, se uma hipotética trajetória de F_1 de um monotongo é quase estacionária, o primeiro coeficiente terá um valor proporcional ao valor médio de F_1 e o segundo coeficiente terá um valor próximo a zero. Em comparação, se a trajetória tem uma inclinação pronunciada, que se poderia esperar no caso de um ditongo como [aĩ] em português brasileiro, o segundo coeficiente terá um valor distante de zero, que pode ser positivo ou negativo, dependendo da direção do movimento. Outros coeficientes quantificam a presença de componentes mais complexos da função-base, como um ciclo completo do cosseno, por exemplo. Seguindo Watson e Harrington (1999), limitamos a análise aos dois primeiros coeficientes de cada formante, uma vez que o estudo dos autores mostrou que eles são suficientes para caracterizar a dinâmica da trajetória formântica.¹⁴ Em nosso caso, aplicamos a DCT aos valores das dimensões avanço e altura das trajetórias médias tanto dos monotongos quanto dos DC, gerando no processo dois valores de coeficiente para cada dimensão. Usamos a função *dct* da biblioteca *emuR* desenvolvida para o ambiente de computação estatística R para realizar essa análise.

3.3. Análise estatística

Para os testes de hipótese envolvendo a variável dependente duração, o teste de Fligner-Killeen (FLIGNER; KILLEEN, 1976) foi aplicado para verificar o pressuposto de homocedasticidade antes da aplicação do teste de comparação de médias. Testes paramétricos foram usados nos casos em que a amostra é homocedástica: teste-*t* no caso de variáveis independentes com apenas dois níveis e análise de variância quando há mais do que dois. Testes não-paramétricos foram usados quando o requisito da homocedasticidade não se verificou: teste de Wilcoxon em lugar do teste-*t* e teste de Kruskal-Wallis em lugar da análise de variância. Quando, nos casos em que a variável independente apresenta mais do que dois níveis, foi preciso fazer comparações múltiplas, recorreu-se a testes-*t* ou Wilcoxon pareados e os valores de *p* foram corrigidos usando a técnica de Benjamini e Hochberg (1995), que controla a quantidade de hipóteses nulas erroneamente descartadas, a chamada *false discovery rate*. Adotamos um nível de 5% para a rejeição da hipótese nula em todos os testes. Usamos o ambiente de computação estatística R em sua versão mais recente no momento da escrita do trabalho para realizar as análises estatísticas, gerar os gráficos e sumários de estatística descritiva.

¹⁴ Em nossos dados, os dois primeiros coeficientes de cada formante foram suficientes para distinguir as trajetórias formânticas dos dois tipos de segmentos analisados, tornando dispensável a extração dos coeficientes superiores ao segundo para os nossos fins.

4. Resultados e discussão

As Figuras 1 e 2 mostram, respectivamente, a trajetória formântica média das vogais cuja realização foi classificada auditivamente como DC (Figura 1) e monotongos (Figura 2). A comparação das duas figuras mostra que há grandes diferenças visuais entre as trajetórias das vogais dos dois tipos. As realizações monotongadas não apresentam trajetórias estritamente estáveis (se isso acontecesse, todos os pontos estariam sobrepostos e não haveria propriamente uma trajetória), porém os movimentos têm amplitude menor quando comparados aos observados nas versões ditongadas. Mais adiante quantificaremos essas diferenças usando a técnica da transformada discreta do cosseno.

Figura 1. Trajetória formântica das instâncias classificadas como DC no espaço definido pelas dimensões do avanço e da altura. O triângulo indica o ponto de início e o quadrado o ponto final de cada trajetória. Os símbolos que representam as vogais estão localizados no ponto mediano da trajetória das versões monotongadas para servir como referência. O símbolo “x” indica o centro do polígono vocálico.

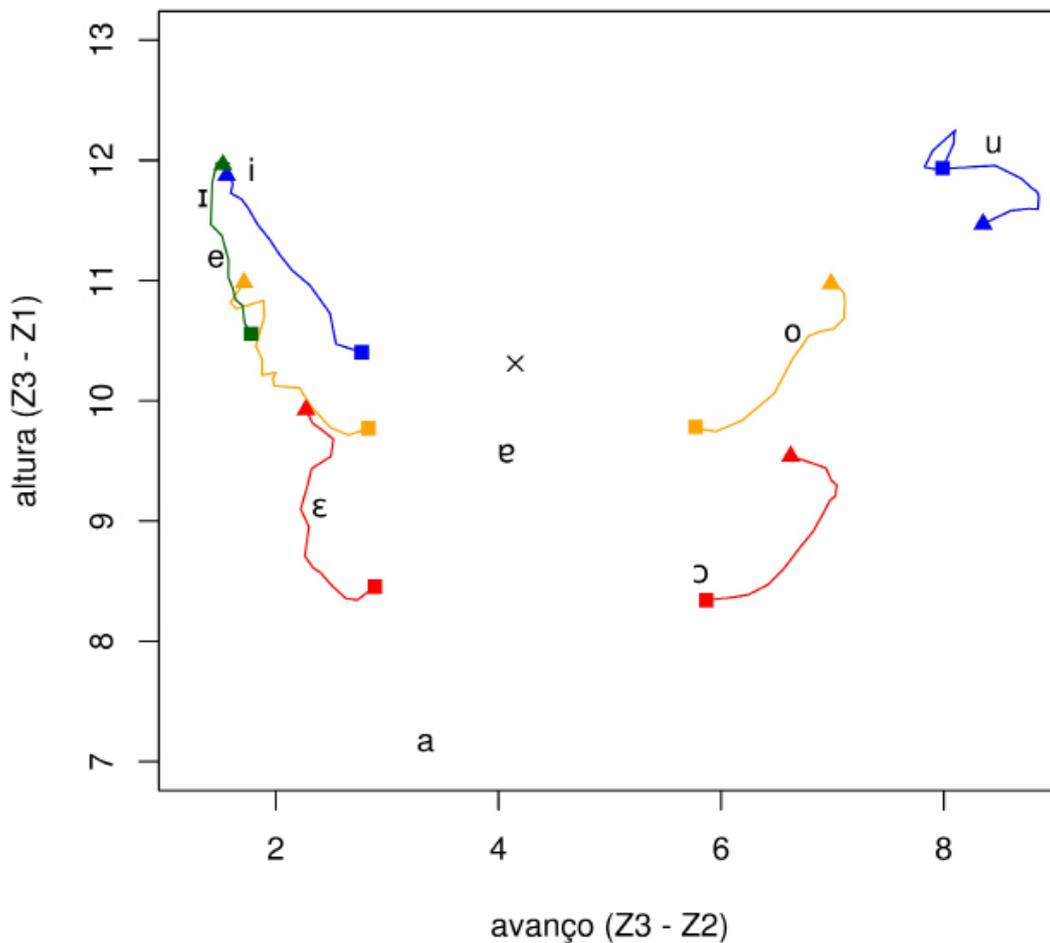
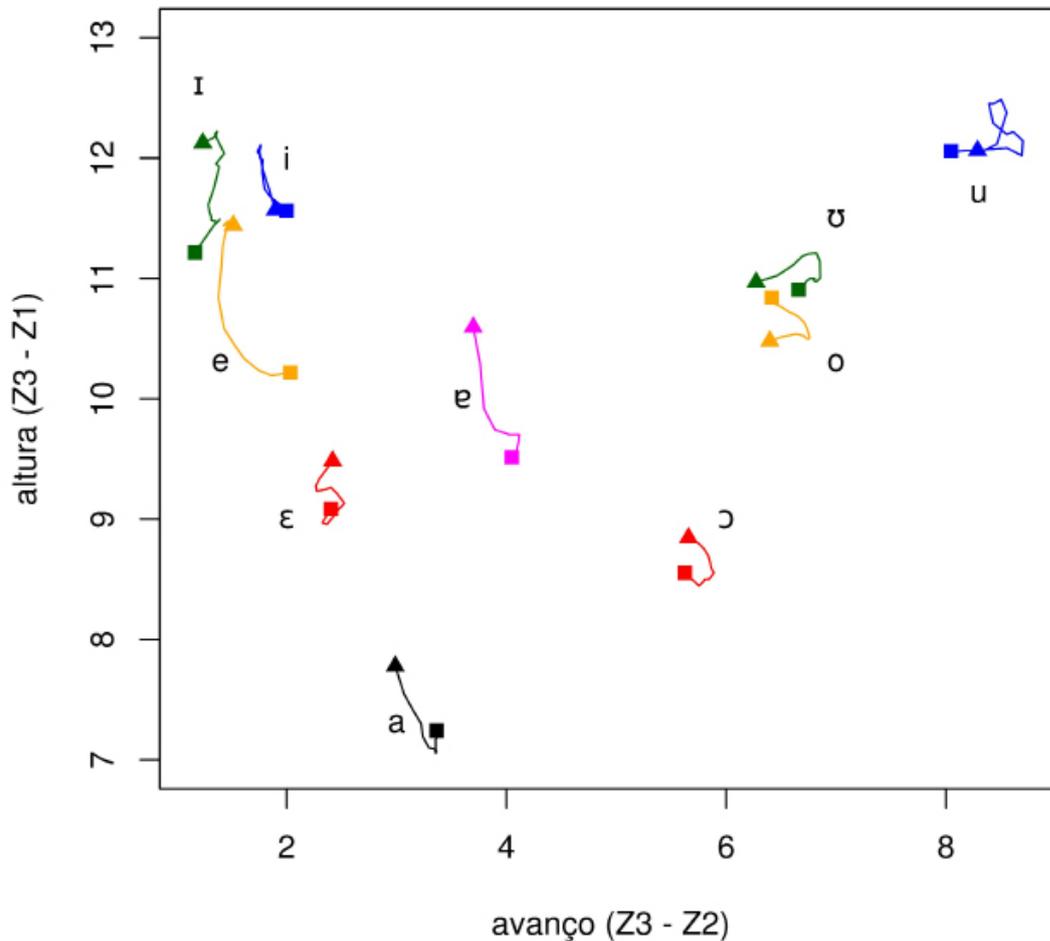


Figura 2: Trajetória formântica das instâncias classificadas como monotongos no espaço definido pelas dimensões do avanço e da altura. O triângulo indica o ponto de início e o quadrado o ponto final de cada trajetória.



A observação, na Figura 1, das trajetórias das vogais classificadas como realizações ditongadas sugere que a transcrição dessas instâncias como ditongos centralizantes é correta, isto é, essas vogais parecem ter um segundo alvo articulatorio na região central do espaço vocálico (ver seção 2).¹⁵ A figura mostra que o formato da trajetória de todas as vogais é compatível com a percepção de um ditongo, dado o movimento de grande magnitude nas duas dimensões, avanço e altura. Em todos os casos, com exceção do /u/, observa-se algum grau de abaixamento e centralização entre o início e o fim da trajetória.¹⁶

¹⁵ Pode-se debater a escolha do símbolo fonético mais apropriado (inclusive se ele deve ser o mesmo em todos os casos), mas o fato de as trajetórias convergirem para o centro do espaço vocálico está fora de dúvida.

¹⁶ Como apontado na seção 3.2, a extração dos formantes de algumas instâncias da vogal /u/ apresentou dificuldades, em função da proximidade entre F_1 e F_2 .

Podemos observar na Figura 1 que o ponto final da trajetória das vogais anteriores /ε, e, i/ na dimensão do avanço é praticamente o mesmo. Entre as vogais posteriores, o ponto final da trajetória das vogais /ɔ, o/ também é praticamente o mesmo na dimensão do avanço, indicando graus de centralização parecidos para essas vogais. Nota-se, também, que as trajetórias não convergem para um mesmo ponto na região central, elas terminam em patamares espaçados que mais ou menos espelham a distância que separa as diferentes vogais no início das trajetórias. Essa é uma diferença marcante entre os presentes resultados e aqueles reportados por Battisti e Oliveira (2014, p. 16) para a variedade de Porto Alegre, que sugerem uma convergência da trajetória de todas as realizações centralizantes para uma região próxima à da vogal representada por [ɐ], muito embora a trajetória representada seja esquemática, não gerada a partir de valores formânticos reais.

4.1. Análise quantitativa das trajetórias formânticas

A Tabela 5 mostra os valores dos dois primeiros coeficientes da transformada discreta do cosseno estimados a partir da trajetória da dimensão da altura (equivalente ao primeiro formante) e do avanço (equivalente ao segundo formante) dos dois tipos de sequência (realização ditongada e monotongada) de sete vogais.

O exame comparativo dos valores do primeiro coeficiente dos dois tipos de vogais mostra que grande parte das diferenças entre os valores do primeiro coeficiente é menor do que 5%. Isso indica que, de forma geral, como seria possível esperar, os dois tipos de vogais ocupam lugares semelhantes no espaço vocálico. A análise do segundo coeficiente, que quantifica a presença de inclinação na trajetória dos dois primeiros formantes, mostra que os DC apresentam valores mais elevados quando comparados às vogais com realização monotongada. Os coeficientes das vogais do primeiro tipo são, em média, 6,3 vezes maiores na dimensão do avanço e 3,7 vezes maiores na dimensão da altura, indicando maior movimentação na dimensão do avanço no caso das realizações com DC. Esses resultados permitem dizer que o grau de movimentação da trajetória é o principal fator que diferencia os dois tipos de realização. As versões monotongadas das vogais [e ɪ] apresentam valores para o segundo coeficiente que não são tão menores do que os valores das versões ditongadas. Um segundo exame auditivo, feito após a análise quantitativa, confirmou a classificação dessas ocorrências como realizações monotongadas e um exame visual da trajetória dos formantes sugere que o movimento na dimensão da altura que essas vogais apresentam é resultado do efeito coarticulatório com a consoante seguinte,

que frequentemente¹⁷ tem um ponto de articulação velar (ou, mais precisamente, pós-velar; ver nota 10).

Tabela 5. Valor do 1º e 2º coeficientes da DCT das trajetórias formânticas das vogais realizadas como ditongos centralizantes ou monotongos

vogal	tipo	dimensão	coeficientes	
			1º	2º
ɛ	ditongo centralizante	avanço	3,44	0,14
		altura	12,82	0,78
	monotongo	avanço	3,37	-0,03
		altura	12,95	0,17
e	ditongo centralizante	avanço	2,90	-0,44
		altura	14,58	0,55
	monotongo	avanço	2,18	-0,16
		altura	15,49	0,70
i	ditongo centralizante	avanço	2,74	-0,51
		altura	16,09	0,68
	monotongo	avanço	2,56	-0,04
		altura	16,79	0,10
I	ditongo centralizante	avanço	2,22	-0,13
		altura	16,03	0,73
	monotongo	avanço	1,88	0,02
		altura	16,63	0,44
ɔ	ditongo centralizante	avanço	9,46	0,38
		altura	12,70	0,58
	monotongo	avanço	9,36	0,06
		altura	15,01	-0,16
O	ditongo centralizante	avanço	9,43	0,59
		altura	14,76	0,61
	monotongo	avanço	9,36	0,06
		altura	15,01	-0,16
u	ditongo centralizante	avanço	11,86	0,42
		altura	16,77	-0,31
	monotongo	avanço	11,75	0,19
		altura	17,25	0,00

¹⁷ Nos nossos dados, esta situação se deu em 8 ocorrências (6 no caso de /e/ e 2 no caso de /i/ pós-tônico).

4.2. Análise da duração das vogais

A Tabela 6 mostra a duração normalizada média das realizações vocálicas classificadas como ditongação carioca e monotongação em posição proeminente e não proeminente no nível frasal. Os resultados estatísticos são apresentados a seguir.

Tabela 6. Duração normalizada (e desvio-padrão) das vogais, expressa em z-scores, em função do tipo de realização do segmento (segunda coluna) e do cruzamento de seu grau de proeminência e do tipo de realização (quinta coluna). A coluna “N” indica o número de instâncias de cada tipo analisadas.

tipo	duração	proeminência	N	duração
ditongo centralizante	3,6 (3,8)	não	17	2,1 (3,7)
		sim	21	4,8 (3,4)
monotongo	0,1 (3,3)	não	65	-0,7 (2,4)
		sim	21	2,7 (4,2)

Os dados da segunda coluna da Tabela 6 mostram que a média de duração normalizada das vogais percebidas como ditongos centralizantes é maior do que a dos monotongos. A realização centralizante apresenta também um nível de variabilidade um pouco mais elevado em comparação à realização monotongada. O efeito de alongamento provocado pelo movimento articulatório centralizante é significativo do ponto de vista estatístico tanto no caso da média [$\chi^2(1) = 31 p < 0,001$] quanto da variância [$\chi^2(1) = 4 p < 0,05$]. O fato da vogal estar ou não em posição de proeminência de nível frasal pode ter um papel na duração da vogal, independentemente de sua realização monotongada ou ditongada por conta do efeito de alongamento em posição de proeminência frasal. No presente *corpus*, as vogais em posição de proeminência têm média de duração normalizada de 3,8 (DP 3,9) e as que não estão em posição proeminente têm média de -0,1 (DP 2,9). Os números mostram que as vogais em posição proeminente sofrem mais alongamento e são também mais variáveis. Esse efeito é significativo do ponto de vista estatístico para as médias [$\chi^2(1) = 31 p < 0,001$] e para a variância [$\chi^2(1) = 14 p < 0,001$]. Os resultados indicam que as duas variáveis independentes, tipo de realização das vogais e proeminência, têm efeito sobre a duração dos vocoides, quer sua realização seja monotongada ou centralizante. É importante, agora, verificar como se dá a interação entre elas, isto é, se o alongamento devido ao tipo de realização e a proeminência da vogal se compõem ou se anulam. A quinta coluna da Tabela 6 sugere que os efeitos se reforçam, pois, entre as vogais em posição proeminente frasalmente, aquelas realizadas como DC são mais longas do que os monotongos (DC

4,8 vs. monot. 2,7; $p < 0,04$). O maior alongamento das vogais realizadas como DC em comparação com os monotongos também se mantém nas posições não proeminentes frasalmente (DC 2,1 vs. monot. -0,7; $p < 0,001$). Crucialmente, o maior alongamento dos DC em comparação aos monotongos não contradiz o efeito de maior alongamento devido à proeminência prosódica, que se mantém quando são considerados os dois tipos de realização das vogais. Entre os DC, a média dos proeminentes é 4,8 e a dos não proeminentes 2,1 ($p < 0,01$). Entre os monotongos, a média dos proeminentes é 2,7 e a dos não proeminentes é -0,7 ($p < 0,001$).

A Tabela 7 mostra o número de ocorrências, a duração acústica e a duração normalizada média das sete vogais para as quais há ocorrências dos dois tipos de sequências (realizações monotongadas e ditongadas). O exame visual mostra que, independentemente da qualidade da vogal, aquelas classificadas como DC tem duração mais longa do que a versão percebida como monotongo.

Tabela 7. Duração bruta (em milissegundos) e normalizada (z-scores) média das diferentes vogais em função do tipo de realização

		ε	e	i	ɪ	ɔ	o	u
ditongo centralizante	N	8	6	7	1	13	2	1
	dur (ms)	194	180	147	260	153	156	167
	dur (z)	4,8	3,7	3,2	14,6	2,5	2,2	4,6
monotongo	N	14	8	19	1	6	10	4
	dur (ms)	119	148	87	145	79	109	96
	dur (z)	0,4	2	0	6,4	-2,2	-0,1	-0,5

Como dissemos na seção 3.1.2, tomamos a decisão de analisar separadamente as vogais que ocorreram em trechos de fala fluente e disfluente. No *corpus*, as vogais a serem analisadas que ocorreram em trechos disfluente correspondem a ocorrências de hesitação, condição em que as vogais frequentemente sofreram alongamento, e de marcadores conversacionais, especialmente a partícula “né”. Segundo a observação de um parecerista, é possível questionar se a normalização da duração das vogais nos trechos de fala disfluente poderia ser feita usando valores de referência que foram obtidos a partir de fala fluente, pois se trata de regimes de fala diferentes. Em razão de nosso desconhecimento de fontes publicadas nas quais se encontrem valores de referência da duração de fones do PB na fala disfluente, usamos os da fala fluente. Para que fiquem claros os possíveis vieses que a normalização pode introduzir neste caso, reportaremos os resultados da estatística

descritiva e dos testes de hipótese tanto para os valores normalizados quanto brutos.

Na amostra analisada neste estudo, há 22 ocorrências de vogais em trechos de disfluência. Destas, 14 foram classificadas como realizações de DC e sua duração bruta média é de 333 ms (DP 209) e normalizada média de 13,3 z-scores (DP 12,3), e oito como realizações monotongadas, com duração bruta média de 315 ms (DP 195) e normalizada média de 11 z-scores (DP 9,9). A comparação das médias indica que o alongamento das vogais em trechos disfluentes é maior e mais variável do que nos trechos fluentes. O teste de homogeneidade de variância indica que o grau de variabilidade dos DC e dos monotongos não é significativamente diferente, quer os dados de duração sejam brutos [$\chi^2(1) = 0,03$ ns] ou normalizados [$\chi^2(1) = 0,04$ ns]. Testes-*t* de duas amostras indicam que a média de duração dos DC e dos monotongos não é estatisticamente diferente quer os dados considerados sejam brutos [$t(20) = 0,02$ ns] ou normalizados [$t(20) = 0,4$ ns].

Uma observação adicional que podemos fazer sobre os dados diz respeito ao fato de que a distribuição de ocorrências de DC e de monotongos não é uniforme nos diferentes contextos. Começando pelos trechos de fala disfluente, nos quais as vogais tendem a sofrer alongamentos consideráveis, das 22 ocorrências quase 65% das vogais foram percebidas como DC. Nos trechos de fala fluente, 66% das 124 ocorrências correspondem a vogais em posições não proeminentes no nível frasal. Nesse conjunto, apenas 26% das vogais foram percebidas como DC. Nas posições percebidas como proeminentes ao nível frasal, grupo que perfaz os 33% restantes do total da amostra, 50% das vogais foram percebidas como DC. Esses dados podem ser interpretados como indicação de que os contextos que estão associados a um alongamento maior da vogal (sílabas proeminentes no nível frasal e trechos de disfluência) são mais propícios para a realização de vogais percebidas como DC.

5. Considerações finais

Neste trabalho, procuramos contribuir para a caracterização acústica do que chamamos de ditongação centralizante na fala carioca, que é um fenômeno característico de uma das variedades mais reconhecíveis do PB e que, até o presente, foi alvo de um número insuficiente de pesquisas fonéticas e fonológicas. Uma novidade introduzida pelo trabalho é a análise da trajetória formântica das sequências de interesse por meio da técnica da normalização temporal da trajetória formântica. A técnica da normalização temporal permite tanto uma análise qualitativa (visual) quanto quantitativa da trajetória dos dois formantes no espaço vocálico, resultando em uma análise mais detalhada do fenômeno.

Os trabalhos anteriores sobre o tema do *ingliding* no PB (BATTISTI; OLIVEIRA, 2014, 2016) analisaram, além da duração, apenas a variação de F_2 , medida em dois pontos: no início e no fim da vogal.

Os dados analisados aqui demonstram que instâncias vocálicas percebidas como DC podem, de fato, ser descritas como a realização de uma vogal com uma finalização articulada na região central do espaço articulatório com variação na altura. As evidências para essa afirmação vêm da observação comparativa das trajetórias de DC e monotongos no espaço vocálico definido pelas dimensões de avanço e altura por um processo de normalização dos formantes e da análise dos coeficientes gerados pela aplicação da transformada discreta do cosseno. Essas duas análises, em conjunto, permitem afirmar que, nos casos de vogais percebidas com DC, as trajetórias apresentam mais movimentação do que aquelas percebidas como monotongos. A maior duração das vogais percebidas como DC também pode ser interpretada como corroboração para sua caracterização como um *inglide*, já que o movimento da língua em direção ao centro do espaço vocálico toma um tempo adicional.

Uma segunda questão importante a respeito do fenômeno dos DC é o seu estatuto linguístico. Não há dúvida de que, no presente momento da variedade carioca, os falantes não fazem uso distintivo dos DC, isto é, a oposição monotongo-DC não é produtiva, ou seja, a comutação de um segmento por outro não gera contraste entre itens da língua. A questão, portanto, é se estamos diante de uma variação livre típica ou se é possível apontar evidências para uma distribuição complementar. Battisti e Oliveira (2016), no seu estudo sobre a variedade falada em Porto Alegre, sugerem que, naquela região metropolitana, o aparecimento de um ou outro alofone é interpretado como índice de estilo, revestido de significados sociais próprios daquela comunidade falante. Não encontramos na literatura nenhuma pesquisa análoga à dos autores mencionados voltada para a variedade carioca. Fica a cargo de trabalhos futuros, que façam uso de metodologia semelhante à empregada por Battisti e Oliveira (op. cit.), levantar hipóteses semelhantes no caso da variedade carioca. Os resultados, por outro lado, também não podem concluir de forma definitiva a respeito do caráter complementar da alofonia entre monotongos e DC. Há indícios, apresentados na seção 4.2, que apontam, malgrado a limitação imposta pelo tamanho relativamente reduzido do *corpus*, para uma preferência dos DC por contextos nos quais se pode esperar o alongamento de segmentos e sílabas, como a posição de proeminência ao nível da frase e trechos de disfluência. Essa preferência, que não é categórica no presente momento, pode vir a ser reinterpretada pelos falantes, ou por um conjunto deles, em algum

momento no futuro, como uma relação necessária e constituir a base para uma regra de distribuição complementar entre alofones monotongados e DC. Os dados apresentados na seção 2 ilustram como outras línguas fizeram o processo de incorporação de ditongos centralizantes em sua fonologia a partir de vogais longas ou alongadas.

De todo modo, nos parece importante esboçar uma explicação para o fato de que dois contextos nos quais vogais sofrem com frequência alongamentos (proeminência prosódica ao nível da frase e hesitações/marcadores discursivos) possam gerar uma variação fonética: um alongamento simples, sem implicações para a articulação da vogal, e um alongamento que vem acompanhado de mudanças articulatórias, como a realização da articulação centralizante documentada no presente trabalho, que é inusual em PB¹⁸ e não é usada nessa língua com valor distintivo no nível lexical.

Para o esboço de explicação, assumimos o modelo da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1986, 1989) com as modificações propostas por Tatham (1995), que incluem um mecanismo de supervisão do processo de produção da fala, que simplifica alguns aspectos das pautas gestuais propostas no âmbito do modelo ao dar conta de detalhes de implementação fonética que o autor classifica como não automáticos e que podem ser explorados de maneira sistemática pelas línguas (para alguns exemplos, vide TATHAM, 1979).

No caso que analisamos aqui, o processo de supervisão deveria garantir que aquilo que foi especificado na pauta gestual como uma vogal simples, um monotongo, seja assim realizado. Para isso, a pauta deve especificar para a variável do trato ‘corpo de língua’ (TB, de *tongue body* em inglês) um lugar e um grau de constrição apropriados para atingir o timbre vocálico intencionado pelo falante. Além disso, deve haver coincidência entre o gesto de TB e o gesto glotal (GLO), com especificação *fechada*, responsável pelo vozeamento. O mecanismo de supervisão fica responsável por garantir que, durante o processo de transformação dessa pauta gestual em especificações de trajetória (operação que é feita pelo componente de dinâmica da tarefa), os gestos de TB e GLO sejam ativados em simultaneidade e durem o tempo determinado pela interação entre a pauta gestual do item lexical e o modelo de duração prosódica (para essa relação, ver BARBOSA, 2006).

Nossa hipótese para explicar o surgimento da articulação centralizante é que essa sequência fônica emerge quando o gesto TB é desativado fora de sincronia relativamente ao gesto GLO por conta de uma falha momentânea no mecanismo de supervisão. A

¹⁸ Contudo, em uma perspectiva tipológica, o fenômeno de formação de ditongos centralizantes constitui uma mudança fonética natural (alguns precedentes foram examinados na seção 2 referentes a línguas que não o PB).

desativação de TB resulta em instruções motoras para o retorno do corpo de língua para a posição articulatória neutra (na qual o dorso da língua está próxima ao centro do espaço vocálico) enquanto o gesto GLO, ainda ativado e em posição *fechada*, produz voz. Essa conformação gera o efeito acústico de *inglide*, observado nas trajetórias da Figura 1. Quanto mais alongada a vogal estiver, mais longa será a trajetória em direção ao centro gerada pela assincronia na desativação dos dois gestos. Voltando aos dados apresentados ao final da seção 4.2, vemos que a maior prevalência relativa da articulação centralizante acontece no contexto de fala disfluente, que é aquele associado aos maiores índices de alongamento, seguido do contexto em que a vogal está em posição proeminente no nível frasal, este também associado a níveis mais elevados de alongamento.

Se aceitarmos a hipótese delineada nos parágrafos acima, devemos concluir que a articulação centralizante que documentamos neste trabalho na verdade não configura um ditongo no sentido de ser uma combinação de vogal e semivogal. O que teríamos, nesse caso, é a especificação de uma vogal simples que é realizada com um *inglide* por conta da desativação do gesto de corpo de língua fora de sincronia com o gesto glotal.

Abreviações

CBN	Central Brasileira de Notícias	ms	milissegundo
DC	ditongos centralizantes	<i>N</i>	quantidade
DCT	transformada discreta do cosseno	PB	português brasileiro
DP	desvio padrão	RS	Rio Grande do Sul
F_i	<i>i</i> -ésimo formante	TB	corpo de língua
GLO	glotal (<i>gesto</i>)		
LCPL	local de constrição da ponta de língua		

Referências

- ALBANO, E. C. *O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado de Letras, 2001.
- ALMEIDA, M. M. S. *Vogais do falar ribeirinho cuiabano*. 2009. 141 f. Tese (Livre-Docência em Filologia e Língua Portuguesa) — Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.
- ARAGÃO, M. do S. S. de. Ditongação e monotongação nas capitais brasileiras. In: HORA, D.; PEDROSA, J. L. R.; LUCENA, R. M. de (Org.). *Anais da ALFAL 2014: XVII Congresso Internacional Asociación de Lingüística y Filología de América Latina (ALFAL)*. João Pessoa: Ideia, 2014. p. 2089-2101.

- ARAÚJO, M. F. R. de. A monotongação do ditongo decrescente [ẽ] no português caxiense. *Revista Estudos Lingüísticos*, Belo Horizonte, v. 8, n. 2, p. 23-51, jul./dez. 1999.
- ASTON, J. A. D.; CHIOU, J. M.; EVANS, J. P. Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Oxford, v. 59, n. 2, p. 297-317, 2010.
- BALODE, L.; HOLVOET, A. The Latvian language and its dialects. In: DAHL, Ö.; KOPTJEVSKAJA-TAMM, M. (Ed.). *Circum-Baltic Languages. Volume I: Past and present*. Amsterdam; Philadelphia: John Benjamins, 2001a. p. 3-40.
- _____. The Lithuanian language and its dialects. In: DAHL, Ö.; KOPTJEVSKAJA-TAMM, M. (Ed.). *Circum-Baltic Languages. Volume I: Past and present*. Amsterdam; Philadelphia: John Benjamins, 2001b. p. 41-79.
- BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes, 2006.
- BATTISTI, E. Realizações variáveis de vogais tônicas em Porto Alegre (RS): Ditongação ou ingliding? *Fragmentum*, Santa Maria, v. 39, p. 60-78, 2013.
- BATTISTI, E.; OLIVEIRA, S. G. de. Alongamento e *ingliding* de vogais em sílabas tônicas no português falado em Porto Alegre (RS). *Revista (Con)Textos Lingüísticos*, Vitória, v. 8, n. 11, p. 39-56, 2014.
- _____. Significados sociais do *ingliding* de vogais tônicas no português falado em Porto Alegre (RS). *Todas as Letras*, São Paulo, v. 18, n. 2, p. 14-29, maio/ago. 2016.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Oxford, v. 57, n. 1, p. 289-300, 1995.
- BISOL, L. Ditongos derivados. *D.E.L.T.A.*, São Paulo, v. 10, p. 123-140, 1994.
- BOERSMA, P.; WEENINK. *Praat: doing phonetics by computer* [Programa de computador]. Versão 6.0.33. Disponível em: <http://www.praat.org/>. Acesso em: 26 set. 2017.
- BROWMAN, C. P.; GOLDSTEIN, L. Towards an articulatory phonology. In: C., Ewan; ANDERSON, J. (Org.). *Phonology Yearbook 3*. Cambridge: Cambridge University Press, 1986, p. 219-252.
- _____. Articulatory gestures as phonological units. *Phonology*, Cambridge, v. 6, 201-251, 1989.
- CABREIRA, S. H. A monotongação dos ditongos orais decrescentes em Curitiba, Florianópolis e Porto Alegre. 115 f. Dissertação (Mestrado em Linguística) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1996.
- CAGLIARI, L. C. *An experimental study of nasality with particular reference to Brazilian Portuguese*. 320 f. Tese (PhD) — University of Edinburgh, Edimburgo, 1977.
- CALLOU, D.; MORAES, J.; LEITE, Y. Apagamento do R final no dialeto carioca: um estudo em tempo aparente e em tempo real. *D.E.L.T.A.*, vol. 14, 61-72, 1998.

- CÂMARA JR., J. M. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes, 1970.
- COLLEY, M. *Diphthongization in Brazilian Portuguese*. 2009. 151 f. Tese (PhD) — Rice University, Houston, 2009.
- _____. Variação fonética nas vogais nasais e nasalizadas no português carioca. In: *Anais do 7º Congresso Internacional da Associação Brasileira de Linguística*. Curitiba: Editora da UFPR, 2011.
- DUARTE, M. E.; PAIVA, M. C. A variação linguística e o papel dos fatores linguísticos. *Revista da Abralin / Associação Brasileira de Linguística*, Curitiba, v. 1 (eletrônico), n. 1 (especial, 1ª parte 2011), p. 91-120, 2011.
- FLIGNER, M. A.; KILLEEN, T. J. Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, Alexandria, v. 71, n. 353, p. 210-213, 1976.
- GILLES, P.; TROUVAIN, J. Luxembourgish. *Journal of the International Phonetic Association*, Cambridge, v. 43, n. 1, p. 67-74, 2013.
- GRABE, E.; KOCHANSKI, G.; COLEMAN, J. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, Berlin, v. 50, n. 3, 281-310, 2007.
- HAKULINEN, A.; VILKUNA, M.; KORHONEN, R.; KOIVISTO, V.; HEINONEN, T. R.; ALHO, I. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura, 2004.
- ITKONEN, T. Välikatsaus suomen kielen juuriin. *Virittäjä*, Helsinki, v. 87, n. 3, p. 349-386, 1983.
- KOHLER, K. J. ToBIG and PROLAB: Two prosodic transcription systems for German compared. *Report at Workshop on Prosodic Labelling*. ICPhS Stockholm. 1995.
- LASS, R. Intradiphthongal Dependencies. In: ANDERSON, J.; DURAND, J. (Ed.). *Explorations in Dependency Phonology*. Dordrecht: Foris Publications Holland, 1987. p. 109-131.
- LEÃO, T. M. R. *Falares regionais: panorama da redução de ditongos decrescentes no Brasil*. 2013. 33 f. Trabalho de conclusão de curso (Licenciatura em Letras) — Departamento de Linguística, Português e Línguas Clássicas, Universidade de Brasília, Brasília, 2013.
- LEITE, Y.; CALLOU, D.; MORAES, J. Processos em curso no português do Brasil: a ditongação. In: HORA, D.; COLLISCHONN, G. (Org.). *Teoria linguística: fonologia e outros temas*. João Pessoa: Universitária, 2003. p. 232-250.
- LEVITT, H.; NEUMAN, A. C. Evaluation of orthogonal polynomial compression. *The Journal of the Acoustical Society of America*, Melville, v. 90, n. 1, 241-252, jul. 1991.
- LOBANOV, B. M. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, Melville, v. 49, n. 2B, 606-608, fev. 1971.
- LOPES, R. A realização variável dos ditongos [ow] e [ej] no português falado em Altamira/PA. 97 f. Dissertação (Pós-Graduação em Letras/Mestrado em Linguística) — Centro de Letras e Artes, Universidade Federal do Pará, Belém, 2002.

LÓPEZ, B. S. *The sound pattern of Brazilian Portuguese (Cariocan dialect)*. 265 f. Tese (Doutorado em Linguística) - University of California in Los Angeles, Los Angeles, 1979.

MOLLICA, M. C. *A influência da fala na alfabetização*. Rio de Janeiro: Tempo brasileiro, 1998.

MORAES, J. A.; WETZELS, L. Sobre a duração dos segmentos vocálicos nasais e nasalizados em português. Um exercício de fonologia experimental. *Cadernos de Estudos Linguísticos*, Campinas, n. 23, p. 153-166, jul/dez. 1992.

MOTA, J. Variação entre /ei/ e /e/ em Sergipe. In: FERREIRA, C. et al. (Ed.). *Diversidade do português no Brasil: estudos de dialetologia rural e outros*. Salvador: Centro Editorial e Didático da UFBA, 1988. p. 143-148.

MUDRAK, O. (Resenha.) Irina Nikolaeva. A Historical Dictionary of Yukaghir. 2006. *Journal of Language Relationship*, Moscou, n. 3, p. 164-173, 2010.

NAU, N. *A short grammar of Latgalian*. München: Lincom Europa, 2011. (Languages of the World/Materials 482.)

NEAREY, T. M. *Phonetic Feature Systems for Vowels*. Bloomington: Indiana University Linguistics Club, 1978.

OSTENDORF, M.; PRICE, P. J., SHATTUCK-HUFNAGEL, S. The Boston University Radio News Corpus. *Boston University Technical Report* No. ECS-95-001. 2001.

OUSHIRO, L.; MENDES, R. B. Sali[ẽ]cia social e mudança linguística: a ditongação de /e/ nasal no português paulistano. *Revista do GEL*, São Paulo, v. 11, n. 32, p. 9-46, 2014.

PAIVA, M. C. Supressão das semivogais nos ditongos decrescentes. In: OLIVEIRA E SILVA, G.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996. p. 217-238.

_____. O percurso da monotongação de [ey]: observações no tempo real. In: PAIVA, M. da C. de; DUARTE, M. E. L. (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Editora Contracapa, 2003. p. 31-46.

PEREIRA, G. *A monotongação dos ditongos <ej>, <ow> e <aj> no português falado em Tubarão (SC): Estudo de casos*. 134 f. Dissertação (Mestrado em Ciência de Linguagem) — Universidade do Sul de Santa Catarina, Tubarão, 2004.

RECTOR, M. *A linguagem da juventude: um estudo geo-sociolingüístico*. Petrópolis: Editora Vozes, 1975.

ROCHA, M. S.; SILVA, A. I. dos A.; NEVES, F. J. L. das. Uma análise sobre a ditongação das vogais tônicas finais seguidas de /s/. *EFDeportes.com*, Buenos Aires, v. 20, n. 208, sep. 2015.

ROSEN, K. M. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics*, Los Angeles, v. 33, n. 4, 411-426, 2005.

- SAPIR, Y. Elfdalian, the Vernacular of Övdaln. In: NYSTRÖM, G. (Ed.). *Fuost konferensn um övdalskų / Första konferensen om älvdalska*. Uppsala: Uppsala Universitet, 2005.
- SCHMALSTIEG, W. R. Primitive east Baltic *-uo-, *-ie- and the 2nd Sg. ending. *Lingua*, v. 10, p. 369-374, 1961.
- SEARA, I. C.; NUNES, V. G.; LAZZAROTTO-VOLCÃO, C. *Fonética e Fonologia do Português Brasileiro*. 2. ed. Florianópolis: LLV/CCE/UFSC, 2011.
- SILVA, A. dos R. Variação fonética em capitais brasileiras: a ditongação diante de /s/ e as realizações fonéticas do /s/ em coda. In: HORA, D. da; PEDROSA, J. L. R.; LUCENA, R. M. de (Org.). *Anais da ALFAL 2014: XVII Congresso Internacional Asociación de Lingüística y Filología de América Latina (ALFAL)*. João Pessoa: Ideia, 2014. p. 4813-4829.
- SILVA, F. de S. O processo de monotongação em João Pessoa. In: HORA, D. da (Org.). *Estudos sociolinguísticos: perfil de uma comunidade*. João Pessoa: CNPq/ILAPPEC/VALPB, 2004. p. 29-44.
- STEVENS, K. N. *Acoustic phonetics*. Cambridge: MIT Press, 1998.
- SYRDAL, A. K.; GOPAL, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, v. 79, p. 1086-1100, 1986.
- TATHAM, M. Some Problems in Phonetic Theory. In: HOLLIEN, H.; HOLLIEN, P. (Org.). *Current Issues in the Phonetic Sciences*. Amsterdam: John Benjamins, 1979, p. 93-106.
- _____. Dynamic articulatory phonology and the supervision of speech production. In: ELENIUS, K.; BRANDERUD, P. (Org.). *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS 95)*. Stockholm: [s.n.], 1995, v. 1, p. 58-61.
- TENIŠEV, Ä. R.; DYBO, A. V. (Ed.). *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov. Prattjurkskij jazyk-osnova. Kartina mira prattjurkskogo ètnosa po dannym jazyka*. Moscou: Nauka, 2006.
- THOMAS, E. R. *Sociophonetics: an introduction*. London: Palgrave Macmillan, 2011.
- TIERSMA, P. The Nature of Phonological Representation: Evidence from Breaking in Frisian. *Journal of Linguistics*, Cambridge, v. 19, n. 1, p. 59-78, março 1983.
- XU, Y. 2007-2015. FormantPro.praat. Disponível em: <http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro/>.
- XU, Y.; GAO, H. FormantPro as a tool for speech analysis and segmentation. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 26, n. 4, p. 1435-1454, 2018.
- VEADO, Rosa M. A redução do ditongo - uma variável sociolinguística. *Ensaios de Linguística*, Belo Horizonte, n. 9, p. 209-229, 1983.

WATSON, C. I.; HARRINGTON, J. Acoustic evidence for dynamic formant trajectories in Australian English Vowels. *The Journal of the Acoustical Society of America*, Melville, v. 106, p. 458-468, 1999.

WETZELS, W. L. The lexical representation of nasality in Brazilian Portuguese. *Probus*, v. 9, n. 2, p. 203-232, 1997.

WETZELS, W. L. Contrastive and allophonic properties of Brazilian Portuguese. In: WANNER, D.; KIBBEE, D. A. (Ed.). *New Analyses in Romance Linguistics: Selected papers from the Linguistic Symposium on Romance Languages XVIII*, Urbana-Champaign, April 7-9, 1988. Amsterdam: John Benjamins, 1991. p. 77-100. (Current Issues in Linguistic Theory 69.)



Data de submissão: 13/03/2018

Data de aceite: 20/11/2018

A VOZ COMO OBJETO DE UMA ANTROPOLOGIA DA ENUNCIÇÃO

VOICE AS AN OBJECT OF AN ANTHROPOLOGY OF ENUNCIATION

Valdir do Nascimento Flores | [Lattes](#) | valdirnf@yahoo.com.br
Universidade Federal do Rio Grande do Sul | CNPq

Resumo: Este texto esboça uma abordagem da voz nos estudos da linguagem, em especial, no interior de uma antropologia da linguagem. O trabalho define a enunciação como uma função que caracteriza fundamentalmente o homem na sua condição de *homo loquens*. Com vistas ao estudo dessa função, é apresentada, em seus aspectos gerais, a ideia de uma antropologia da enunciação, que estuda o saber sobre o homem que advém de sua enunciação. Defende-se, em suma, uma perspectiva teórica que trata dos efeitos da presença da língua no homem. Finalmente, ilustra-se o proposto com uma análise de comentários feitos acerca da voz de cantores brasileiros.

Palavras-chave: Antropologia da enunciação; Voz; Enunciação; Falante.

Abstract: This text outlines an approach to the voice in language studies, especially within an anthropology of language. The study defines enunciation as a function that fundamentally characterizes men in their condition of *homo loquens*. Aiming to study this function, the idea of an anthropology of enunciation is presented in its general aspects. Such anthropology studies the knowledge about men that comes from their enunciation. Therefore, we defend a theoretical perspective that deals with the effects of the presence of language in men. Finally, we illustrate the proposal with an analysis of comments made about the voice of Brazilian singers.

Keywords: Anthropology of enunciation; Voice; Enunciation; Speaker.

Introdução

O objetivo geral deste artigo – de certa forma, explícito em seu título – é esboçar uma proposta de entendimento da voz no escopo de uma antropologia da enunciação. Assim formulado, tal objetivo pode, dada a aparente clareza que o caracteriza, encobrir duas necessidades de explicitação conceitual: a noção de “voz”¹ mobilizada e o enten-

¹ Utiliza-se o destaque das aspas na palavra, na Introdução e no decorrer do texto, sempre que se tratar de *menção* ao termo e não de *uso* da palavra.

dimento do que vem a ser uma “antropologia da enunciação”. Nenhum desses termos é evidente no campo dos estudos da linguagem, merecendo, portanto, maior atenção.

Desse ponto de vista, então, poder-se-ia dizer que há um objetivo anterior ao apresentado acima, cujo duplo propósito seria, de um lado, explicitar os termos pelos quais se pode conceber a “voz” como um objeto passível de abordagem no âmbito dos estudos da linguagem e, de outro lado, evidenciar as linhas fundamentais de uma visada sobre a linguagem, concebida como uma antropologia da enunciação, que comporta a “voz” como objeto de estudo. Em vista disso, este texto se organiza em três partes.

A primeira (cf. item 1) busca definir o sentido de “voz” mobilizado neste estudo. Faz-se isso a partir da análise histórico-crítica da maneira como a “voz” comparece nos estudos da linguagem, em geral, e na linguística, em particular. Conclui-se esse item formulando algumas razões que impuseram uma espécie de “mutismo” da voz nos estudos da linguagem. Ao finalizar a primeira parte, ainda se concebem duas possibilidades – antagônicas metodologicamente – de abordagem da voz: vinculando-a à língua, entendida como *factum grammaticae* (MILNER, 1995), ou vinculando-a ao falante, entendido como *homo loquens*, em que a propriedade *loquens* define uma função que caracteriza fundamentalmente o homem.²

A segunda parte (cf. item 2) embasa, em linhas gerais, o que se entende pela expressão “antropologia da enunciação”, perspectiva que coloca o falante no centro dos estudos da linguagem e que vê na capacidade *loquens* do homem a fonte de um saber de natureza linguística. A antropologia da enunciação, nesses termos, é apresentada como o estudo do saber do homem que advém da sua capacidade de enunciar. Nessa parte, também se fundamentam os princípios metodológicos de uma tomada antropológica da enunciação, considerando, em especial, a proposição do “contorno de sentido”, visível na função metalinguística natural de usar a língua para comentar a língua, como uma categoria de análise da antropologia da enunciação.

Finalmente, na terceira parte (cf. item 3), já de posse das considerações advindas dos itens anteriores, busca-se apresentar a voz no escopo teórico-metodológico esboçado. Ilustra-se o proposto com uma análise de comentários feitos acerca da voz de cantores e cantoras brasileiros.

² Por recomendação dos avaliadores “ad hoc” do artigo, destaca-se que o uso do significante “homem”, na construção teórica aqui buscada – de inspiração benvenistiana – circunscreve-se à ideia de “humano”. Não se ignora, no entanto, que o mesmo significante, quando usado em outros contextos, pode implicar discursos sexistas.

1 A voz humana, esta desconhecida

A ideia principal deste item está contida em seu título: a voz humana é desconhecida tanto dos estudos da linguagem, em geral, quanto dos estudos da linguística, em particular.

Inicialmente, cabe observar que a expressão “voz humana” – que não por acaso dá nome à célebre peça de Jean Cocteau (*La voix humaine*), cuja força poética resulta da profundidade dos sentimentos, *demasiado* humanos, da personagem – não tem sentido óbvio, sem ambiguidade. Uma rápida vista de olhos no campo dos estudos gerais sobre a linguagem permite ver que à aparente evidência da expressão “voz humana” é possível contrapor algumas considerações que, de um lado, problematizam sua configuração fenomenológica e, de outro lado, colocam em xeque sua real abordagem pelos estudos da linguagem (filosóficos, linguísticos etc.). Observem-se alguns exemplos.

Herman Parret, semiótico e filósofo da linguagem, propõe uma discussão nos seguintes termos: “há *sons, ruídos* produzidos pelo ser humano, que não pressuponham a voz, que sejam, em certo sentido, *anteriores à voz*, como o balbúcio das crianças e o riso?” Ao que responde: “Tudo dependerá evidentemente da definição de voz que se precisará construir, na sua relação com a sonoridade e com o ruído”. Acrescenta, ainda, que a palavra “voz” pode ser usada apenas metaforicamente, quando aplicada a outros seres (com vida ou não) que não o homem (animais ou máquinas). Isso, no entanto, não é suficiente para “isolar, entre os ‘ruídos do mundo’, uma classe específica de ruídos ou de sons gerados pela voz humana” (PARRET, 2002, p. 23, *itálicos no original, tradução minha*).

Grosso modo, o que Parret parece questionar é a existência de um critério seguro que permita isolar ruídos, em geral, de sons produzidos pela voz, em particular. Com base em Aristóteles, ele conclui: “a voz humana, ao contrário do ruído e da voz animal, é eufônica e sinfônica” (PARRET, 2002, p. 26, *tradução minha*), quer dizer, ela é *semantizada e estetizada*.

Não muito longe do que diz o semiótico, mas indo em outra direção, o filósofo italiano Giorgio Agamben pergunta:

Existe uma voz humana, uma voz que seja voz do homem como o fretenir é a voz da cigarra ou o zurro é a voz do jumento? E, caso exista, é esta voz linguagem? Qual a relação entre voz e linguagem, entre *phoné* e *logos*? E se algo como a voz humana não existe, em que sentido o homem pode ainda ser definido como o vivente que possui linguagem? (AGAMBEN, 2008, p. 10, *itálicos no original*).

Distanciando-se de uma visão aristotélica, que articula a voz em termos da relação entre *phoné* e *logos*, Agamben supõe um hiato entre voz e linguagem que, no constructo teórico do filósofo, pode abrir espaço para uma ética, na justa medida em que é “um espaço vazio” entre a potência do dizer e o ato.

Também o medievalista Paul Zumthor em resposta à pergunta de André Beaudet, que indagava “como se explica que nós ainda não tenhamos uma ciência da voz?” (ZUMTHOR, 2005, p. 62), assim se manifesta:

A voz se destaca destas evidências por assim dizer tão óbvias, que a reflexão científica só alcança alguns de seus aspectos. Uma ciência da voz deveria abarcar tanto uma fonética quanto uma fonologia, chegar até uma psicologia da profundidade, uma antropologia e uma história. (ZUMTHOR, 2005, p. 62).

E conclui mais adiante: “a língua é mediatizada, levada pela voz. Mas a voz ultrapassa a língua; ela é mais ampla do que ela, mais rica [...] a voz, utilizando a linguagem para dizer alguma coisa, se diz a si própria, se coloca como presença” (ZUMTHOR, 2005, p. 63).

Finalmente, Corrado Bologna, talvez a primeira e mais reiterada personalidade dos estudos sobre voz, sentencia:

A sua natureza é essencialmente física, corpórea; está relacionada com a vida e com a morte, com a respiração e com o sono; emana dos mesmos órgãos que presidem à alimentação e à sobrevivência. Antes mesmo de ser o suporte e o canal de transmissão das palavras através da linguagem, a voz é um imperioso grito de presença. (BOLOGNA, 1987, p. 58).

O que essas ínfimas passagens de trabalhos tão densos ilustram? Simplesmente que o tema da “voz” está longe de ser considerado algo pacífico entre aqueles que problematizam a existência humana; daí o pouco poder autoexplicativo da expressão “voz humana” e sua quase nenhuma presença no campo dos estudos da linguagem.

A voz, de fato, tem tantas faces, tantos aspectos, tantas possibilidades de ser escutada que pode, facilmente, interessar a muitos campos do conhecimento, ou a nenhum. No entanto, e isso é inegável, a voz não prescinde da linguagem, embora não equivalha *pari passu* a ela. E, sendo a linguística uma autoridade em matéria de linguagem, não é fácil não ceder à tentação de ver a voz no escopo da linguística. Mas não de qualquer linguística: de uma linguística que inclua o *homo loquens* em seu horizonte de possibilidades.

1.1 Os termos do mutismo da voz no campo da linguística

A filósofa Adriana Cavarero, em um livro de rara beleza, *Vozes plurais - Filosofia da expressão vocal*, parece encontrar a origem tanto da natureza da problematização, na atualidade, quanto da ausência da voz nos estudos da filosofia (e, também, da linguística). Segundo ela, o *logos* perdeu a voz. Quer dizer, desde a sua origem – e Cavarero está pensando especialmente na filosofia platônica –, a “filosofia tapa os ouvidos” e promove uma desvocalização do *logos*. O *logos*, a razão, o discurso, a linguagem, enfim, dizem respeito, na tradição do pensamento filosófico, à conexão de palavras. É nesse plano da conexão, que “liga” de acordo com determinadas regras, que está centrada a atenção da filosofia: “Centrada, inclusive com prejuízo – mas talvez fosse melhor dizer: sobretudo com prejuízo – do plano acústico da palavra” (CAVARERO, 2011, p. 50-51).

Cavarero tem razão. A ausência da voz coincide com a (excessiva?) presença do *logos*, da razão, que abafa a voz. E a que se deve esse “ensurdecimento racional”?

Ora, a filósofa considera que a voz é uma “unicidade que faz de cada um de nós um ser diferente dos demais” (CAVARERO, 2011, p. 17), e que tal unicidade “não é característica do homem em geral, mas de cada ser humano na medida em que vive e respira” (CAVARERO, 2011, p. 18). No entanto, adverte: não se trata de uma relação tautológica entre a própria voz e o próprio ouvido, “mas de um comunicar-se da unicidade que é, ao mesmo tempo, uma *relação* com outra unicidade” (CAVARERO, 2011, p. 20, itálico no original), isto é, a voz implica a escuta. Assim, “no registro da voz, ecoa a condição humana da unicidade [...] [e] tal condição é essencialmente relacional” (CAVARERO, 2011, p. 22-23). Ao estudo da problemática da unicidade relacional da voz, Cavarero (2011) chama de *fenomenologia vocálica da unicidade*.

Assim entendida a voz – como um dado indiscutivelmente singular e único – é fácil compreender o ensurdecimento que a filosofia tem em relação a ela:

A singularidade inimitável de cada ser humano, a unicidade encarnada que distingue cada um de qualquer outro, é, para os gostos universalistas da filosofia, um dado supérfluo, se não desconcertante, além de epistemologicamente impróprio. (CAVARERO, 2011, p. 24).

É uma evidência: esse ensurdecimento gera um distanciamento, pois a tradição filosófica e epistemológica determina que, em proveito da abstração generalizada de vocação universalista da ciência, se evite tratar o que é único, singular e irreproduzível, do que a voz é um exemplo.

Tal análise e suas consequências são estendidas à linguística³ pela própria Cavarero, o que merece ser citado na íntegra apesar da extensão:

A despeito de sua aceleração novecentista e de seus complexos desenvolvimentos, pode-se de fato desenterrar o percurso teórico que leva a linguística moderna a herdar do estatuto metafísico uma desatenção programática sobre a unicidade da voz. O que, obviamente, não depõe contra a legitimidade dos estudos linguísticos nem sequer diminui o seu interesse, **mas testemunha, de modo exemplar, como os saberes dirigidos aos fenômenos da palavra podem ocupar-se da voz sem preocupar-se com a singularidade de cada voz.** Em outros termos, a voz – estudada na perspectiva da linguagem e, ainda mais, numa perspectiva que entende linguagem como sistema – torna-se a esfera geral das articulações sonoras na qual a unicidade do som é, paradoxalmente, aquilo que não *soa*. A linguagem enquanto código, a sua alma semântica que aspira ao universal, torna imperceptível, na voz, o *próprio* da voz. A unicidade plural das vozes não passa pelo filtro metodológico do ouvido linguístico. (CAVARERO, 2011, p. 24-25, itálicos no original, negritos meus).

A análise acima pode causar algum espanto junto aos linguistas, que acreditam estudar o que é tangível da voz, em seus tratados de fonética e fonologia.

Porém, Cavarero pode ter alguma razão. Na fonética e na fonologia, realmente, nada há que fale da singularidade da voz, ao menos não nos termos explicitados pela filósofa. No máximo, a linguística abriga, nos limites bastante rígidos de suas metodologias, aspectos implicados na prosódia – em seu viés acústico ou fonológico (SCARPA, 1999) –, na produção da fala – a partir de recursos tecnológicos de análise acústica (MARCHAL; REIS, 2012) –, no tratamento fonético *stricto sensu* (CAGLIARI, 2009), entre outros. O *Dicionário de fonética e fonologia*, por exemplo, não reserva nenhuma de suas entradas a um verbete sobre “voz”, embora, evidentemente, trate com bastante propriedade todos os aspectos “relevantes para o estudo do português e da fonética e da fonologia em geral” (SILVA, 2011, p. 8).

É bem verdade, entretanto, que, na atualidade, especificamente com relação aos estudos prosódicos, encontra-se alguma preocupação em tratar algo que se aproxime do que Cavarero considera próprio à unicidade da voz. Barbosa (2010, p. 1), por exemplo, diz que “[...] no cenário de pesquisa atual, a prosódia tem seu campo de estudo nos domínios linguístico, paralinguístico e extralinguístico”. Nos três domínios, segundo o autor,

³ Deve-se excetuar o excelente *La vive voix: essais de psycho-phonétique*, de Ivan Fónagi (1991).

[...] estudam-se as funções prosódicas de demarcação (indicadores de constituintes prosódicos, como sílabas, palavras fonológicas, grupos acentuais, sintagmas entoacionais, entre outros), proeminência (saliência de um constituinte prosódico em relação a outro) e de marcação discursiva (marcadores de turno num diálogo, modalidade da frase, entre outros). (BARBOSA, 2010, p. 1).

Essas funções prosódicas estão ligadas à entoação, ao ritmo, à “[...] imbricação entre restrições biomecânicas ligadas à produção da fala [...] e restrições linguísticas e paralinguísticas ligadas à percepção da fala [...]” (BARBOSA, 2010, p. 1). Para o autor, as funções linguísticas do ritmo, da entoação, da acentuação, do acento lexical e frasal têm lugar nos estudos linguísticos, já os fenômenos “[...] languageiros e comunicativos [...]” (BARBOSA, 2010, p. 1) – marcadores discursivos, ênfase, atitudes, emoções e os fenômenos ligados a fatores sociais e individuais (como gênero e sexo, classe social, faixa etária, condição de saúde) – têm lugar nos estudos extralinguísticos e paralinguísticos.

Admitida a distinção feita por Barbosa (2010) entre linguístico, extralinguístico e paralinguístico, o que Cavarero (2011) abriga na ideia de uma *fenomenologia vocálica da unicidade* teria alguma relação com os ditos fenômenos “languageiros e comunicativos”, uma vez que ali seriam abordados “fatores individuais”. Porém, Barbosa (2010) insiste sobre o fato de tais fenômenos terem lugar em estudos extralinguísticos e paralinguísticos, o que resguarda uma especificidade ao “linguístico” que vai ao encontro do que Cavarero (2011, p. 25) denomina de “linguagem como sistema”. Sendo os aspectos individuais considerados extralinguísticos ou paralinguísticos, a linguística pode eximir-se de abordá-los.

Em outras palavras, os estudos linguísticos, quando olham para a voz, tratam-na no interior de uma concepção de linguística, de ciência, que é refratária aos ditos “fatores individuais”. O efeito disso é a exclusão da unicidade da voz do escopo da linguística, apesar da explícita admissão da existência de elementos na voz que não podem ser tratados no mesmo quadro metodológico da fonética e da fonologia. Em suma, o gesto de reconhecimento da vinculação da voz ao falante é exatamente o mesmo que a exclui da linguística.

Há, de certa forma, um aspecto paradoxal na enunciação vocal, singular e individual, com o qual o linguista tem dificuldades de lidar, e é Émile Benveniste quem assinala isso. Em *O aparelho formal da enunciação*, ele considera que “o [aspecto] mais imediatamente perceptível e o mais direto – embora de um modo geral não seja visto em relação ao fenômeno geral da enunciação – é a realização vocal da língua” (BENVENISTE, 1989a, p. 82).

Benveniste (1989a, p. 82) afirma que “os sons emitidos e percebidos [...] procedem sempre de atos individuais, que o linguista surpreende sempre que possível em uma produção nativa, no interior da fala”. No entanto – e é aí que se mostra o aspecto paradoxal da enunciação vocal individual –, normalmente, se procura, em linguística, “[...] eliminar ou atenuar os traços individuais da enunciação fônica recorrendo a sujeitos diferentes e multiplicando os registros, de modo a obter uma imagem média de sons, distintos ou ligados” (BENVENISTE, 1989a, p. 82). Ele adverte:

Mas cada um sabe que, para o mesmo sujeito, os mesmos sons não são jamais reproduzidos exatamente, e que a noção de identidade não é senão aproximativa mesmo quando a experiência é repetida em detalhe. Estas diferenças dizem respeito à diversidade das situações nas quais a enunciação é produzida. (BENVENISTE, 1989a, p. 82-83).

Ou seja, Benveniste é claro em destacar – e resgatar – a materialidade corporal da enunciação, isto é, o falante, além de assinalar com veemência a paradoxal atitude do linguista.

A voz – entendida como única e singular – remete ao falante, ao *homo loquens*, ao corpo, que, como se sabe, excede o limite da ciência. As disciplinas linguísticas, na abordagem que dão ao tema, desviam da realidade do falante em favor da abstração seja da língua como sistema, seja do falante como mero *enunciador, locutor, emissor* etc.

Consequentemente, exclui-se a voz – e, com ela, o falante – do horizonte da linguística. O homem na sua condição de falante, o *homo loquens*, não opera na análise linguística da linguagem nem mesmo quando é da voz que se busca falar (FLORES, 2017). Evitando tratar da voz, daquilo que há de único nela, ou seja, do falante, a linguística evita rivalizar com uma forma de fazer ciência na qual impera a repetição do mesmo como critério de verificação.

A despeito desse cenário adverso à abordagem da voz no campo da linguística, vale ainda indagar: é possível uma linguística que comporte, em algum sentido, o estudo da voz? Uma resposta afirmativa exigiria que a linguística se reconfigurasse epistemologicamente de forma a comportar a singularidade como um fenômeno passível de ser tomado como objeto de investigação, o que, em termos linguísticos, coincidiria com a realidade do falante. O olhar do linguista deveria deixar de mirar a abstração das propriedades linguísticas para fixar os olhos no *homo loquens*. Não mais a língua, mas o falante que fala sob o efeito de ser constituído pela língua.

Poder-se-ia, então, admitir – ou, ao menos, conjecturar – dois pontos de vista de tratamento da voz no campo da linguística.

O primeiro, compatível com a cientificidade que caracteriza fortemente a área, em que, apesar de se reconhecer as especificidades de uma enunciação vocal singular, trata-se no âmbito generalizante dos “fatores individuais” extralinguísticos. Têm lugar, aqui, as abordagens fonéticas e fonológicas cujo filtro metodológico evita o falante na condição de *homo loquens*. Mesmo que possam apresentar matizes de técnicas e procedimentos, os estudos oriundos desse ponto de vista têm em comum o olhar para o que Jean-Claude Milner (1995) chama de *factum grammaticae*:

Se existe uma ciência da linguagem, ela deve atribuir propriedades à linguagem; isso supõe que ela atribui propriedades a cada língua; e isso supõe, por sua vez, que ela atribui propriedades a cada fragmento de uma língua. Mas o que garante que seja simplesmente possível, de maneira geral, atribuir propriedades a dados da língua? A resposta sobre esse ponto é simples: a garantia consiste somente na existência, de fato, do que chamamos gramáticas. Toda e qualquer gramática consiste em atribuir propriedades a um dado de língua. Se as gramáticas existem de fato, então, é preciso concluir que uma tal atribuição é possível. Em outras palavras, a linguística como ciência se apoia sobre o *factum grammaticae* (MILNER, 1995, p. 54-55, tradução minha).

Do *factum grammaticae* – que poderia ser entendido tanto como o sistema que os linguistas tentam estabelecer quanto como as descrições propostas para esse sistema – está excluído o *homo loquens*, o que conduz, imediatamente, ao segundo ponto de vista possível acerca da voz. Neste, o olhar é dirigido menos à “gramática” de uma língua e mais ao sujeito falante, ao *homo loquens*. Desse olhar, aqui nomeado antropologia da enunciação, deve surgir uma perspectiva que coloca em evidência um saber sobre a língua que advém do fato de o homem falar.

A antropologia da enunciação é uma outra linguística, talvez menos científica, que supõe que o fato de a língua ser constitutiva do homem dá-lhe a condição de especial “conhecedor” da língua.

2 A língua no homem: uma antropologia da enunciação

O objetivo desta parte é definir, em linhas gerais, a enunciação como uma *função* que caracteriza o *homo loquens* e identifica o caráter fundamentalmente verbal da condição humana. A enunciação – em uma inspiração nitidamente benvenistiana – é entendida, aqui, como o ato de dizer algo que coloca em cena um saber sobre a natureza *loquens*

do homem. Essa função constitui um objeto antropológico⁴ na justa medida em que dá a conhecer os efeitos da presença da *língua no homem*. A antropologia da enunciação visa, portanto, esse saber sobre o homem que advém do fato de o homem falar.

Com a formulação *a língua no homem*, opera-se, conscientemente, uma inversão à expressão *o homem na língua*, com a qual Émile Benveniste nomeia a quinta parte de seus *Problemas de linguística geral*. *A língua no homem*, colocada no centro de uma antropologia da enunciação, evoca como objeto fenômenos – a voz talvez seja um dos mais evidentes, mas há outros (a tradução, a aquisição, a dissolução, a arte verbal etc.) – que dão as condições necessárias e suficientes para que se investigue um saber sobre a língua a partir do sujeito falante.

Nesses termos, a antropologia da enunciação resgata para o seu interior a grande lição de Jakobson (1974, p. 34), para quem “a linguística interessa-se pela linguagem em todos os seus aspectos – pela linguagem em ato, pela linguagem em evolução, pela linguagem em estado nascente, pela linguagem em dissolução”.

Enfim, que consequências há, para uma linguística, pensar *a língua no homem*?

Em um primeiro momento, é preciso compreender que se trata de uma antropologia no sentido amplo de “conhecimento do homem” e não em algum sentido específico ou aplicado a um domínio qualquer. Além disso, trata-se de uma antropologia da enunciação, quer dizer, do dizer do homem acerca da presença da língua nele. Ou seja, o homem fala do fato de experienciar, em diferentes fenômenos, a presença da língua nele.

Desse ponto de vista, a antropologia da enunciação estuda o fato de o falante tematizar a sua posição de falante ao tratar de fenômenos em que ele está implicado como falante. Assim, dá-se destaque ao retorno reflexivo que o falante produz ao falar sobre como opera a língua nele.

A categoria que dá acesso aos procedimentos analíticos que desvelam a presença da língua no homem é o comentário, entendido como um *contorno de sentido*, isto é, uma operação natural do falante que visa à explicação e à compreensão das formas e da presença da língua nele. O *contorno de sentido* é uma espécie de hermenêutica natural, na medida em que o falante fala para atribuir sentido à sua posição de falante em função de um dado

⁴ A reflexão antropológica visada aqui – de inspiração kantiana (LALANDE, 1996, p. 73) – diz respeito ao homem em geral. Ou também, como lembra Todorov, a respeito da expressão “antropologia geral”: “a palavra pode também ser tomada em seu sentido literal de ‘conhecimento do homem’ para designar o conceito que temos acerca do humano e que estaria subjacente às diversas explorações das ciências humanas assim como aos discursos morais ou políticos, ou, ainda, à filosofia” (TODOROV, 2014, p. 9). O linguista Émile Benveniste também utiliza o termo neste sentido: “vemos todo o conjunto das ciências humanas se desenvolver, formar-se toda uma grande antropologia (no sentido de ‘ciência geral do homem’)” (BENVENISTE, 1989c, p. 38).

fenômeno linguístico. O *contorno de sentido* pertence, portanto, à função metalinguística natural e produz uma hermenêutica natural.

De certa maneira, o comentário, o *contorno de sentido*, é uma narrativa que o falante produz sobre sua história de falante, o que o alça à condição de um etnógrafo da própria língua (FLORES, 2016). Em resumo, o *contorno de sentido* é um comentário que o falante faz sobre a experiência linguística – dele ou de um outro falante – no contexto de um fenômeno linguístico qualquer. O *comentador* enfoca o conjunto dos meios expressivos utilizados por ele mesmo ou pelo outro; sobre esse conjunto é produzido um *contorno de sentido*, ou seja, um saber que o falante articula, uma interpretação sobre um elemento qualquer de um dado fenômeno,⁵ Fazendo uso da natureza metalinguística da linguagem, o falante coloca em prática a capacidade de delimitar mecanismos necessários e suficientes para que a propriedade da linguagem de interpretar a si mesma (BENVENISTE, 1989b, p. 62) se efetive.

Ao linguista interessado no que se está propondo, caberia ver as condições de elaboração de um pensamento sobre a língua cujo parâmetro é o falante. Esse pensamento, sob a aparência de um comentário “leigo”, esconde uma sofisticada operação, qual seja: a de cada falante voltar reflexivamente sobre as possibilidades significativas particulares que cada fenômeno de língua opera.

Em outras palavras, o falante, ao falar sobre os efeitos que os diferentes fenômenos de língua têm sobre ele, se historiciza como *homo loquens* em sua língua. Nesse sentido, uma antropologia da enunciação não estuda dados, mas fenômenos e, em especial, o que o falante diz de sua relação com esses fenômenos.

No entanto, é bom que se advirta, a antropologia da enunciação, ao mirar o que diz o falante sobre fenômenos da língua – a tradução, a aquisição, a dissolução e a voz, entre outros –, não faz teoria deles. Isto é, a antropologia da enunciação, ao estudar o que diz o falante de sua experiência com a diversidade das línguas em uma situação de tradução, por exemplo, não produz uma teoria da tradução, mas uma reflexão sobre o efeito da tradução na experiência de falante. É nesse sentido que se estuda um saber que advém do fato de o homem enunciar; é nesse sentido, também, que se considera que é *a língua no homem* que determina esse saber.

Ocorre o mesmo com a voz. A antropologia da enunciação, ao tematizar a presença da voz no homem, não produz uma teoria da voz, uma fonética ou uma fonologia novas,

⁵ Com essa ideia, espera-se deixar claro que o falante pode comentar tanto a sua própria experiência de falante como a de outro qualquer. Nos dois casos, se trata de uma “etnografia” de si, uma vez que o falante explicita sua relação tanto na condição de falante propriamente dito como na condição de quem produz uma escuta.

mas tão-somente uma reflexão sobre o homem como falante, ou melhor, sobre a experiência do homem acerca do fato de que tem voz.

Talvez, enfim, a partir do olhar de uma antropologia da enunciação, a linguística esteja em melhores condições de responder a excelente questão de Claude Hagège (1985, p. 10): “que lugar cabe à linguagem na definição do homem”?

3 A voz no contexto de uma antropologia da enunciação

Em 2012, a revista *Rolling Stones Brasil*, de circulação nacional, especializada em música, publicou, em sua 73ª edição, uma lista do que considerou, à época, “as 100 maiores vozes da música brasileira” (AS 100, 2012). Na apresentação da lista, assim se manifesta a revista:

Na edição de aniversário de seis anos da *Rolling Stone Brasil*, mantemos a tradição de apresentarmos as listas definitivas da música brasileira e mundial. Desta vez, o ranking, elaborado por um time de 60 especialistas, define quais são as 100 maiores vozes da história da nossa música. **O que é ter uma voz marcante?** Naturalmente, qualidades como potências e afinação são bem-vindas. Mas muitas vozes transmitem emoção e expressividade, mesmo sendo fora do padrão. (AS 100, 2012, grifo meu).

Independentemente da pertinência, ou não, da elaboração de listas como essas – que hierarquizam elementos de naturezas muito distintas e com características tão singulares que inviabilizam qualquer termo de comparação –, o fato é que a revista faz acompanhar a lista uma pergunta, grifada acima, fundamental para os propósitos deste texto. Observar os termos pelos quais a revista encaminha respostas a essa pergunta deve permitir reunir elementos para pensar os aspectos antropológicos da enunciação ligados à voz.⁶

Além disso, a revista publicou, juntamente com a lista, uma série de comentários feita por também cantores (inclusive alguns constantes no dito inventário) acerca das vozes então “classificadas”. Tais comentários merecem alguma atenção no âmago de um estudo antropológico da enunciação.⁷ Observe-se, abaixo, a lista de comentários relativos aos “dez primeiros colocados”, em ordem crescente:

- 1) O cantor Seu Jorge a propósito da voz de Tim Maia: “O timbre de Tim Maia era muito particular e muito pessoal. De grande extensão, ia do grave ao agu-

⁶ Há autores que se dedicam exclusivamente ao estudo da “palavra cantada”, tema este que não recobre a noção de “fala”. Para maiores esclarecimentos ver: Cagliari (2009, p. 159) e Carmo Jr. (2007 e 2012).

⁷ Observe-se que, no tratamento do corpus, a seguir, são analisados comentários de terceiros sobre as vozes dos cantores destacados. Isso se coaduna com o que foi dito na nota 4 (acima) quanto à possibilidade de se desenvolver uma reflexão acerca da atividade metalinguística sobre a voz que considere a atenção reflexiva em relação à voz do outro.

do e era marcante, com o sotaque da música brasileira. Você percebe samba, jovem guarda, forró, tudo fundido dentro dessa particularidade de black music” (AS 100, 2012);

- 2) A cantora Maria Rita sobre a voz de Elis Regina: “Buscava a perfeição no que fazia, não nivelava por baixo, não se contentava com pouco. E ela era, sim, uma musicista, com destaque especial pra sua noção de divisão, onde e como colocar notas curtas, onde sustentar uma nota, sempre pensando na história que a letra contava” (AS 100, 2012);
- 3) A cantora Rita Lee a propósito da voz de Ney Matogrosso: “[...] é incomparável. Nunca houve entre os cantores brasileiros uma figura tão sedutora, chique e atrevida. Públicos feminino e masculino são hipnotizados por sua voz e presença. Não há quem não fique apaixonado. O timbre da voz é inigualável. Quando se ouve, sabe-se na hora quem é. Isto se chama personalidade e ele a aplica em todos os poros de sua arte. Bobagem falar em momentos mais importantes da vida de um artista do calibre dele, que nunca se repete. Foi inesquecível quando o vi pela primeira vez. Foi uma aparição do outro mundo: um ET elegante vestindo um kabuki muito louco com uma voz assexuada, cantando uma ciranda portuguesa” (AS 100, 2012);
- 4) O cantor Max de Castro a propósito da voz de Wilson Simonal: “Simon al inaugurou uma nova escola de canto no Brasil. Ele uniu todas as escolas vocais, desde o cool da bossa nova até a potência vocal, acrescentando uma influência do suingue, na maneira mais criativa de se interpretar uma música. Não somente por saber cantar as notas originais, mas também por criar uma divisão diferente e novas possibilidades de melodias paralelas” (AS 100, 2012);
- 5) O cantor Caetano Veloso a propósito da voz de Maria Bethânia: “Bethânia surgiu para o público como apenas e exclusivamente uma voz. Foi na Bahia. O diretor de teatro Álvaro Guimarães montou *O Boca de Ouro*, de Nelson Rodrigues, e o espetáculo abria com um longo blackout em que se ouvia a voz de uma garota desconhecida cantando ‘Na Cadência do Samba’, de Ataulfo Alves. Ela tinha 17 anos, mas era já essa voz de timbre rico e afeto intenso que nos impressiona até hoje. Ninguém via a figura que sustentava aquela voz nas trevas. Mas o espetáculo começava com uma força ímpar no teatro mundial. Conheço a voz de Bethânia desde dentro: ela foi se desenvolvendo pertinho

- de mim – e tinha os elementos genéticos que estão presentes na minha própria voz, na de meus outros irmãos, na de meus filhos” (AS 100, 2012);
- 6) O cantor Marcelo Jeneci a propósito da voz de Roberto Carlos: “Sua voz é a voz do amigo que a gente precisa ter durante toda a vida. Esse feito fez dele um dos compositores mais importantes na história da música brasileira” (AS 100, 2012);
 - 7) A cantora Tulipa Ruiz a propósito da voz de Gal Costa: “[...] é uma das mais bonitas que já ouvi. O jeito que ela se apropria das letras, da história, me faz pensar que ela está totalmente à vontade dentro de uma música. Gal se relaciona com o que canta” (AS 100, 2012);
 - 8) O cantor Ney Matogrosso a propósito da voz de Caetano Veloso: “[...] se transformou num cantor maravilhoso quando voltou da Inglaterra, deu um pulo enorme. Até então eu o achava em tudo excelente, menos como cantor. Ele descobriu lá um jeito de cantar que depois evoluiu muito” (AS 100, 2012);
 - 9) O cantor Paulinho da Viola a propósito da voz de Clara Nunes: “[...] era uma cantora que a gente ouvia e sabia imediatamente quem era. Tinha uma voz muito bonita e foi muito importante para o samba e para a música brasileira. Como intérprete, fez parte de uma geração de cantoras que você ouvia e sabia imediatamente quem era, porque elas tinham sua forma de interpretar, deixavam sua identidade na voz” (AS 100, 2012);
 - 10) O cantor Samuel Rosa a propósito da voz de Milton Nascimento: “[...] é um dos principais artistas da música brasileira. Ele sintetiza, na inovadora assinatura musical dele, uma série de habilidades. Tem voz linda e um timbre inconfundível e muito particular. Ao ouvi-lo, não se faz referência direta a algo da nossa música, mesmo havendo mistura riquíssima ali de música mineira, sacra, hispânica e latina, e até dos Beatles e do folk de Dylan” (AS 100, 2012).

Os comentários acima, em que a voz é considerada na constituição da música, encontram eco em um belo texto de Roland Barthes, “A música, a voz, a língua”, no qual ele afirma: “a minha avaliação da música passa pela voz” (BARTHES, 2009b, p. 266). E acrescenta, a propósito da voz em sua relação com a música:

A voz humana é, com efeito, o lugar privilegiado (eidético) da diferença: um lugar que escapa a toda a ciência, pois não há nenhuma ciência (fisiologia, história, estética, psicanálise) que esgote a voz: classifiquem, comentem

historicamente, sociologicamente, esteticamente, tecnicamente a **música**, haverá sempre um resto, um suplemento, um *lapsus*, um não dito que se designa ele próprio: a voz. (BARTHES, 2009b, p. 266, itálico no original, negrito meu).

Tem razão Barthes: a voz humana é refratária à indiferença, portanto, refratária a um discurso completamente conformado a um ideal de cientificidade, porque “*não há nenhuma voz humana no mundo que não seja objeto de desejo – ou de repulsa*” (BARTHES, 2009b, p. 266, itálico no original).

Os comentários acima, na medida em que constituem um discurso *sobre* a voz, são um excelente exemplo do que fala Barthes, além de, ao traduzirem a voz pela língua, ser uma ilustração do que foi antes considerado uma hermenêutica natural. É desse ponto de vista que serão tratados aqui.

Como, nesses comentários, a língua traduz a voz? O conjunto dos comentários permite, inicialmente, surpreender vários aspectos da voz – interligados entre si – que buscam ser “colocados em palavras”.

Há, inicialmente, a apreciação a partir da eleição de elementos técnicos – “timbre” (1); “extensão” (1), “grave” (01); “agudo” (1); “notas curtas” (2), “sustentar uma nota” (2); “timbre da voz” (3); “uma divisão diferente” (4); “timbre” (5); “timbre” (10) –; há, também, a apreciação estética sob a forma de uma avaliação predicativa (cf. BARTHES, 2009a; 2009b) – “era marcante” (1); “é incomparável” (3); “é inigualável” (3); “é a voz do amigo” (6); “é uma das mais bonitas” (7); “tinha uma voz muito bonita” (9); “tem voz linda” (10) –; e há, por último, a constatação da *fenomenologia vocálica da unicidade* de que fala Cavarero (2011) – “era muito particular e muito pessoal” (1); “quando se ouve, sabe-se na hora quem é” (3); “a gente ouvia e sabia imediatamente quem era” (9); “identidade na voz” (9); “muito particular” (10).

Esse conjunto – traduzido na língua via terminologia técnica, predicativo e reconhecimento da unicidade – constitui uma primeira possibilidade de a língua falar sobre a música; constitui também uma primeira possibilidade de situar o *homo loquens*.

No entanto, há, também, o que excede essa apreciação e que diz respeito à figura do sujeito falante, do *homo loquens*. Uma pergunta que cabe, aqui, é: como a língua traduz o saber que advém da enunciação, objeto da antropologia da enunciação? Ou ainda: como a voz, nesses comentários, pode ser vista como objeto de uma antropologia da enunciação?

Em uma primeira interpretação, poder-se-ia pensar que o fato de o conjunto dos comentários dirigir-se sempre a um nome próprio, o cantor ou a cantora “donos da voz”,

evidenciaria uma instância de referência ao falante. Não é essa a proposta que se está defendendo aqui.

O *homo loquens*, termo que resume a proposta de uma antropologia da enunciação, não é o *homo sapiens* – o protagonista das histórias sobre a evolução humana –, nem mesmo as representações figurativas que a linguística construiu ao longo de sua abertura à natureza discursiva da linguagem (os locutores, enunciadores, sujeitos etc.). O *homo loquens* é constituído na linguagem e sua natureza é feita de linguagem. Ele é sujeito por ser falante, um sujeito falante. Assim, antes de ser uma categoria da antropologia da enunciação, ele é a sua condição. Por isso, a enunciação do *homo loquens* deixa à mostra um saber sobre a sua natureza, sobre os efeitos da presença da língua nele.

E que saber os comentários deixam vir à tona?

Ora, esse saber aparece precisamente na falta de “objetividade” que gira em torno da palavra técnica, falta esta que é preenchida de maneira a evidenciar um saber sobre a língua que é de ordem quase metafísica: o timbre de Tim Maia, além de suas características técnicas, tem “sotaque da música brasileira” (1); a “noção de divisão” de Elis Regina é acompanhada de um “sempre pensando na história que a letra contava” (2); o “timbre da voz” de Ney Matogrosso hipnotiza (3); o timbre da voz de Maria Betânia é “rico” e de “afeto intenso” (5).

Soma-se a isso um outro ponto fundamental: os aspectos interligados entre si tentam falar sobre algo para o que a língua não oferece uma palavra pronta: a voz de Milton Nascimento (10), por exemplo, além de “linda” e de “um timbre inconfundível e muito particular”, “não faz referência direta a algo da nossa música”, embora apresente uma “mistura riquíssima”. A voz de Maria Bethânia (5), por sua vez, requer até uma metáfora espacial, pois o comentador afirma conhecê-la “desde dentro”.

Não há a palavra certa para nomear a voz. Resta apenas contorná-la de sentido. Esse contorno (FLORES, 2016) permite ao falante falar semanticamente da materialidade da língua, o que explicita um saber, o seu saber, acerca dela. Esse saber cumpre uma função quase etnográfica na economia – no arranjo ou modo de funcionar de diversos elementos de um conjunto maior – do uso da língua. É nesse sentido que se poderia defender que o falante, ao falar da língua – muito especialmente, da materialidade da língua –, é um etnógrafo da própria língua. Essa etnografia se apresenta na interpretação que faz da língua, através dos comentários que a tomam, no caso da voz, via significante.

Enfim, vale lembrar que o que se vê nos comentários é uma tentativa de a língua traduzir a voz através daquilo que se escuta dela. O que pulsa numa voz escapa àquilo que

geralmente costuma se tomar como sua porção “tangível” (o tom, o volume e as tantas tentativas de adjetivá-la – voz rouca, voz áspera etc.). Porém, os sons emitidos pelo homem, através da voz, são dinâmicos e efêmeros. Logo, falar em voz é falar em efeitos. Em que lugar a voz encontra abrigo? Na escuta do outro.

Assim, o próprio da voz humana é o fato de que ela significa de uma maneira paradoxal: ela é, simultaneamente, o específico de cada um – que, inclusive, não admite reprodução ou cópia –, e o partilhado com o outro. A voz tem essa propriedade de ser única e, ao mesmo tempo, social. Talvez, por isso, seja tão própria para dar voz ao *homo loquens* na linguística.

Conclusão

Do que foi dito, uma conclusão se impõe: de um lado, não se pode ignorar que a voz não se conforma à linguística, ao menos não aos moldes de ciência sobre os quais se edifica a linguística moderna; de outro lado, não se pode negar que a voz humana não é algo que se opõe à linguagem.

No entanto, quando se trata da voz, está-se frente ao *homo loquens* – ao *homem falando com outro homem*, como diria Benveniste (1995, p. 285) –, ao homem que se apresenta como tal ao falar. E o homem transcende o objeto da linguística. Seria necessário refundar o olhar da linguística sobre a linguagem para, nele, incluir o homem falante e, com este, a sua voz. Seria preciso, portanto, renunciar à mestria que o discurso científico da linguística fornece.

As conclusões esboçadas neste texto, então, não pertencem ao campo da ciência da demonstração, embora não cheguem a ignorá-lo completamente. Em outras palavras, pela natureza da problemática que a voz evoca, nem seria justo vinculá-la integralmente e nem opô-la completamente à ciência linguística.

A expressão *homo loquens* sintetiza a proposta aqui esboçada e, nesse sentido, pode encontrar abrigo em uma outra linguística, que admite o falante, logo, o homem, no seu interior. Esboça-se aqui, então, uma linguística do falante, neste artigo, também nomeada “antropologia da enunciação”.

Referências

- AGAMBEN, G. *Infância e história: destruição da experiência e origem da história*. Belo Horizonte: Editora da UFMG, 2008.
- AS 100 maiores vozes da música brasileira. *Rolling Stone Brasil*, São Paulo, n. 73, out. 2012. Disponível em: <http://rollingstone.uol.com.br/listas/100-maiores-vozes-da-musica-brasileira/>. Acesso em: 15 jan. 2018.

BARBOSA, P. A. Prosódia: uma entrevista com Plínio A. Barbosa. *ReVEL*, v. 8, n. 15, 2010. Disponível em: http://www.revel.inf.br/files/entrevistas/revel_15_entrevista_plinio.pdf. Acesso em: 15 jan. 2018.

BARTHES, R. O grão da voz. In: _____. *O óbvio e o obtuso*. Lisboa: Edições 70, 2009a. p. 255-264.

_____. A música, a voz, a língua. In: _____. *O óbvio e o obtuso*. Lisboa: Edições 70, 2009b. p. 265-271.

BENVENISTE, E. O aparelho formal da enunciação. In: _____. *Problemas de linguística geral II*. São Paulo: Pontes, 1989a. p. 81-90.

_____. Semiologia da Língua. In: _____. *Problemas de linguística geral II*. São Paulo: Pontes, 1989b. p. 43-67.

_____. Esta linguagem que faz a história. In: _____. *Problemas de linguística geral II*. São Paulo: Pontes, 1989c. p. 29-40.

_____. Da subjetividade na linguagem. In: _____. *Problemas de linguística geral I*. São Paulo: Pontes, 1995. p. 284-293.

BOLOGNA, C. Voz. In: *Enciclopédia Einaudi - Volume 11 - Oral/escrito; argumentação*. Lisboa: Imprensa Nacional Casa da Moeda, 1987.

CAGLIARI, L. C. *Elementos de fonética do português brasileiro*. São Paulo: Paulistana, 2009.

CARMO JR. J. R. *Melodia & Prosódia: um modelo para a interface música-fala com base no estudo comparado do aparelho fonador e dos instrumentos musicais reais e virtuais*. 2007. Tese de doutorado, FFLCH/USP, São Paulo, 2007.

_____. Sobre a gramática da palavra cantada. *Cadernos de estudos linguísticos*, Campinas, v. 54, n. 2, 2012. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636602/4321>. Acesso em: 30 setembro 2018.

CAVARERO, A. *Vozes plurais: filosofia da expressão vocal*. Belo Horizonte: Editora da UFMG, 2011.

FLORES, V. N. O falante como etnógrafo da própria língua: uma antropologia da enunciação. *Letras de Hoje*, v. 50, p. 90-95, 2016.

_____. A voz, “essa cabeça de Medusa”. In: MALISKA, M. E.; SOUZA, P. (Org.). *Abordagens da Voz: a partir da Análise do Discurso e da Psicanálise*. Campinas: Pontes, 2017. p. 119-133.

FÓNAGY, I. *La vive voix: essais de psycho-phonétique*. Paris: Payot, 1991.

HAGÈGE, C. *L'Homme de paroles*. Paris: Librairie Arthème Fayard, 1985.

JAKOBSON, R. *Linguística e comunicação*. São Paulo: Cultrix, 1974.

LALANDE, A. *Vocabulário técnico e crítico da filosofia*. São Paulo: Martins Fontes, 1996.

MARCHAL, A.; REIS, C. *Produção da fala*. Belo Horizonte: Editora da UFMG, 2012.

MILNER, J.-C. *Introduction à une science du langage*. Édition abrégée. Paris: Éditions du Seuil, 1995.

PARRET, H. *La voix et son temps*. Bruxelas: De Boeck & Larcier, 2002.

SILVA, T. C. *Dicionário de fonética e fonologia*. São Paulo: Editora Contexto, 2011.

SCARPA, E. Apresentação. In: _____. (Org.). *Estudos de prosódia*. Campinas: Editora da UNICAMP, 1999. p. 7-15.

TODOROV, T. *A vida em comum: ensaio de antropologia geral*. São Paulo: Editora Unesp, 2014.

ZUMTHOR, P. Presença da voz. In: _____. *Escritura e nomadismo*. São Paulo: Ateliê Editorial, 2005. p. 61-102.



Data de submissão: 21/01/2018

Data de aceite: 13/06/2018

VOCAL ASPECT OF SOCIAL LAUGHTER DURING VIRTUAL INTERACTION

ASPECTOS VOCAIS DO RISO SOCIAL NA INTERAÇÃO VIRTUAL

Takaaki Shochi | [CLLE Curriculum](#) | takaaki.shochi@labri.fr
CLLE-ERSSàB UMR5263 & LaBRI CNRS UMR 5800, Bordeaux, France

Marine Guerry | [CLLE Curriculum](#) | marine.guerry@gmail.com
CLLE-ERSSàB CNRS UMR5263, Bordeaux, France

Hanako Suzuki | [Ritsumeikan University Curriculum](#) | suzukih@fc.ritsumei.ac.jp
Ritsumeikan University, Japan

Mami Kanzaki | [J-Global](#) | kanmami.224@gmail.com
Ritsumeikan Global Innovation Research Organization, Japan

Jean-Luc Rouas | [LaBRI Curriculum](#) | jean-luc.rouas@labri.fr
LaBRI, CNRS UMR 5800, Talence, France

Toyoaki Nishida | [Nishida Lab](#) | nishida@i.kyoto-u.ac.jp
Kyoto University, Japan

Yoshimasa Ohmoto | [J-Global](#) | ohmoto@i.kyoto-u.ac.jp
Kyoto University, Japan

Resumo: O presente artigo enfoca os diversos tipos de riso registrados durante interações sociais reais realizadas em um ambiente de imersão virtual. Neste experimento, investigamos se os seres humanos são capazes de discriminar risos sociais de risos espontâneos e gargalhadas audiovisuais, sem terem acesso a nenhum contexto de produção do riso. Para tal, realizamos dois experimentos perceptuais considerando apenas o estímulo auditivo, por um lado, e o estímulo audiovisual, por outro. Os sujeitos dos testes de percepção são falantes nativos franceses e japoneses. Cada sujeito ouviu e viu 162 risadas e escolheu uma resposta entre três possibilidades: social, espontânea ou irreconhecível. Os resultados de ambos os experimentos mostram que todos os participantes são capazes de discriminar esses dois tipos de riso com um grau de confiança bastante elevado, mesmo

sem as informações contextuais: a taxa de identificação correta para o riso espontâneo é de cerca de 70%, com uma quantidade semelhante de riso social considerando tanto os estímulos visuais quanto os audiovisuais. A seguir, foram extraídas medidas acústicas para cada risada, a fim de investigar possíveis diferenças entre os dois tipos de riso. A análise multifatorial demonstra que os comportamentos perceptuais e algumas características acústicas (F0 e duração) estão correlacionados. Observamos, particularmente, uma diferença significativa entre o riso social e o riso espontâneo através dos traços da duração total e da duração do vozeamento. Por último, realizamos mais um experimento perceptual sobre a subcategorização dos risos sociais com base em três fatores sociais: estado físico do falante, envolvimento do falante e distância psicológica. Os resultados mostram que o riso social é evocado de forma semelhante no que diz respeito ao estado físico do falante e envolvimento do falante. Entretanto, no que diz respeito à distância social, há divergência de percepção entre franceses e japoneses sendo que nestes últimos predomina a percepção da distância psicológica.

Palavras-chave: Riso; Reconhecimento; Análise acústica; Percepção; Cultura.

Abstract: The current paper focuses on the various types of laughter recorded during real social interactions in a virtual immersive environment. With this experiment, we investigate whether human beings are able to discriminate social from spontaneous laughter on the basis of auditory or audiovisual laughs presented outside any context. Towards this aim, we carried out two perceptual experiments proposing audio alone and audiovisual conditions, which were taken by French and Japanese native subjects. Each subject listened to (or looked at) 162 laughs and chose one response among three possibilities: social, spontaneous or unknown. The results of both experiments show that all participants are able to discriminate these two types of laughter with quite good confidence without contextual information: the correct identification rate for spontaneous laughter is about 70% with a similar amount for social laughter in audio alone and audiovisual conditions. We then extracted acoustic characteristics for each laugh in order to investigate potential differences between the two types of laughter. A multifactorial analysis showed that perceptual behaviors and some acoustic features (F0 and duration) are correlated. Especially, we observe a significant difference between social and spontaneous laughter through the features of total duration and voiced duration. Finally, we conducted a perceptual experiment on the subcategorization of social laughs based on three social factors: speaker's

physical state, speaker's involvement and psychological distance. The results show that social laughter is characterized by similar contexts for both groups of listeners, except by Japanese subjects who regard the psychological distance between partners as distant.

Keywords: Human laughter; Recognition; Acoustics; Perception; Culture.

1 Introduction

Comprehensive knowledge about the vocal characteristics of social affective interaction has been neglected for a long time because of the lack of sufficient understanding about the cognitive processing of various affective meanings as well as technical realization of such expressions. However, the automatic recognition and synthetic realization of such affective meaning became one of the important issues for researchers of various scientific research fields like social robotics, medical hearing tools or language learning tools (ISHI et al., 2016; SHOCHI et al., 2016).

Human social interaction consists in an exchange of social information conveyed by voice, eye contact, gestures, facial expressions, sighs or laughter (SHOCHI et al., 2016; WICHMANN, 2000; CAMPBELL et al., 2005). Among these modalities, laughter is probably one of the most important behaviors in the development of speech and in human and animal communication (PROVINE, 1996; PROVINE, 2001; OWREN; BACHOROWSKI, 2007). Laughter is often considered as a physical reaction to external stimuli, which is often linked to positive valence (i.e. joyful reaction). In Jacykiewicz and Ringeval (2014) the authors report that laughter is usually provoked by external stimuli, and organized on three different axes: neuro-hormonal involving periaqueductal gray, the reticular formation with inputs from cortex basal ganglia and from the hypothalamus (WILD et al., 2003), including muscular inputs and the respiratory axis.

According to the field of Gelotology¹, laughter is classified under 15 categories (BERK, 2001). The first three ones: “smirk”, “smile” and “grin” are perfectly voluntarily controllable without a laugh sound. Both “snicker” and “giggle” use facial muscles as “grin” but are accompanied by laugh sounds. “Chuckle” involves chest muscles with deeper pitch, and “chortle” involves muscles of the torso. This type of laugh usually provokes laughter in others (BERK, 2001). When a laugh is accompanied by barking or snorting, it is usually classified in the category of “laugh [which] involves facial and thoracic muscles as well as abdomen and extremities” (BERK, 2001, p. 327). All these eight types of

¹ “Gelotology” from the Greek “γέλως / *gelos*” (meaning “laughter”) and “λογία / *logia*” (study of). (BUTLER, 2005).

laughs are described as controllable and can be voluntary simulated. On the contrary, seven other laughs: “cackle”, “guffaw”, “howl”, “shriek”, “roar”, “convulse” and “die laughing” are characterized by an uncontrollable and involuntary nature. This physical description of laughter is deeply rooted in human biology. Therefore, these characteristics should apply to every human being. However, laughter also serves very important social roles to bring about positive, mutually beneficial relationships among people and communities (ERICKSON et al., 2009; SCOTT et al., 2014). When laughter plays social roles, it must be strongly influenced by the culture we belong to. For instance, R.P. Luís Fróis, a Portuguese missionary who travelled in Asia in the 16th century, mentioned cultural differences in the use of laughter between Portuguese and Japanese in his manuscript “About the difference between Europe and Japan”.

“In our country, a feigned laugh is considered hypocritical; In Japan, it is a mark of elegance and good disposition”

(Luís Fróis, 1585 p. 87)

“In our country, politeness is made of a serene and grave attitude; The Japanese infallibly do so with all sorts of little simulated laughs”

(Luís Fróis, 1585 p. 89)

These descriptions suggest two important pieces of information about laughter. Firstly, laughs are not always linked to amusement: people frequently simulate laughs, which play an important social role in face-to-face communication. Secondly, there is a clear perceptual gap between laughter of two cultures. Such cultural differences influence how people laugh depending on various social contexts.

Indeed, a few studies have already tried to reveal the cultural aspects of laughter, especially focusing on the distinction between two types of laughs: social laughter and an amusing (or mirthful) one (PROVINE, 2001; VETTIN; TODT, 2005; ERICKSON et al., 2009; TANAKA; CAMPBELL, 2011).

In Scott et al. (2014), the authors suggest the existence of two different types of laughter: spontaneous and volitional (or social), distinguished by neuro-physiological differences. Spontaneous laughter is considered an involuntary reaction to external stimuli. It is supposed to be innate because it occurs even before the first words. Physiological changes during such involuntary laughter are quite different from what occurs during a voluntary one. For instance, involuntary laughter is characterized by a higher activation of hypothalamus than for the voluntary one, and the chest expansion and amplitude of sound waves show more regular cycle patterns than the voluntary one, which exhibits a

speech-like pattern. On the other hand, social laughter is supposed to be an intentional communicative act, recruited in order to set up a positive relationship or to tone down a conflictive tension.

Concerning the acoustic realization of these various types of laughter, Trouvain (2003) summarizes an analysis of various types of laughter in three levels: “segmental”, “syllable” and “phrasal”. With regard to acoustic analyses at the phrasal level, Tanaka; Campbell (2011, 2014) made an acoustic distinction between “mirthful” involuntary spontaneous laughs and “politeness” laughs which is closely linked to social relationships; spontaneous mirthful laughs tend to include chuckles and ingressive sounds which were rare in polite laughs. Moreover, mirthful involuntary laughs tend to be longer than polite social laughs. Other recent works (ANIKIN; LIMA, 2017; BACHOROWSKI et al., 2001; KIPPER; TODT, 2001; VETTIN; TODT, 2005) support the idea that spontaneous laughter has a higher and a more variable fundamental frequency (hereby F0), as well as a higher variability in acoustic parameters in general. In addition, spontaneous laughter is also characterized by a longer duration with shorter bursts, ingressive and chuckle sounds (TANAKA; CAMPBELL, 2011; LAVAN et al., 2016). However, there is no significant difference for both types of laughter regarding breathiness and mouth aperture.

According to our hypothesis, (1) human beings are able to discriminate between social (voluntary) laughs and mirthful spontaneous (involuntary) ones using only audio cues without visual or contextual information; (2) perceptually determined mirthful involuntary laughter may have common acoustic cues among different cultures; and on the contrary, (3) volitional social laughter may be perceived differently from one culture to another due to cultural conventions.

Following these hypotheses, the current research investigates (1) whether French and Japanese subjects can discriminate between social volitional laughter and spontaneous involuntary laughter using a) only auditory laughs or b) audiovisual laughs, extracted from an immersive virtual interaction environment, and without any context. Independently we aim at investigating (2) the acoustic characteristics for each type of laughter in French and Japanese. Moreover, (3) we try to describe a socio-linguistic context for a majority of social volitional laughter.

2 Corpus

The stimuli were recorded in an immersive virtual environment at Kyoto University, Japan. This database consists of spontaneous affective speech recorded during a virtual

reality game played by three participants. The game was designed to study communication between people in virtual environments and was made using Unity². The game was designed by two scenario designers. They discussed how to control the virtual characters in the recording scenario for inducing spontaneous affective reactions in players. The motions of the virtual characters were made from interaction motions of humans captured by a Kinect sensor so to provide natural motions. The scenario designers carefully arranged the captured motions and pre-recorded motions, which were distributed by the asset store of Unity3D. Each player was alone in his own individual immersive virtual environment (completely surrounded by displays or in an immersive dome). Inside the immersive environment, participants could look around at the virtual game world with a low cognitive load, as in the real world. In addition, each player could communicate with the others using Skype with cameras and microphones, so they could recognize their facial expressions and hear each other's voices. The Skype windows were displayed on the left side of the player. They were not obstructive when the player interacted with the virtual character. The players were required to communicate in order to solve various tasks instructed by three different virtual characters. One of the main interests of this approach is that each participant can be recorded individually (Figure 1). A total of 12 spontaneous affective speech data files of 9 Japanese (2F/7M) and 3 French (1F/2M) were recorded. All participants were university students recruited at Kyoto, Japan. All Japanese native subjects grew up in Japan (Kansai region), and the 3 native French speakers were foreign students at Kyoto University, but were not able to communicate in Japanese at the moment of the experiment.

A total of 254 sequences containing only laughter were manually segmented using PRAAT (BOERSMA; DAVID, 2013). A first pilot test was conducted in order to investigate which acoustic features distinguish mirthful spontaneous vocalizations of laughs from volitional forms which are considered as social laughs (LAVAN et al., 2016; ANIKIN; LIMA, 2017; JÜRGENS et al., 2011). 7 experimenters (3 Japanese males and 4 French (3F/1M) were instructed to listen and to annotate each sample using two labels: "spontaneous" and "social". According to a selection threshold criterion based on more than 70% of identification of the stimulus perceived as "spontaneous", a set of 27 spontaneous laughs was chosen. These 27 laughs consist in 6 laughs uttered by French speakers and 21 laughs from Japanese speakers. Another pilot test designed to select the 27 volitional social laughs produced by 2 French and 25 Japanese speakers was done under the same criterion as for the spontaneous one by the same experimenters as for the first pilot test.

² Unity is a cross-platform game engine developed by Unity Technologies.

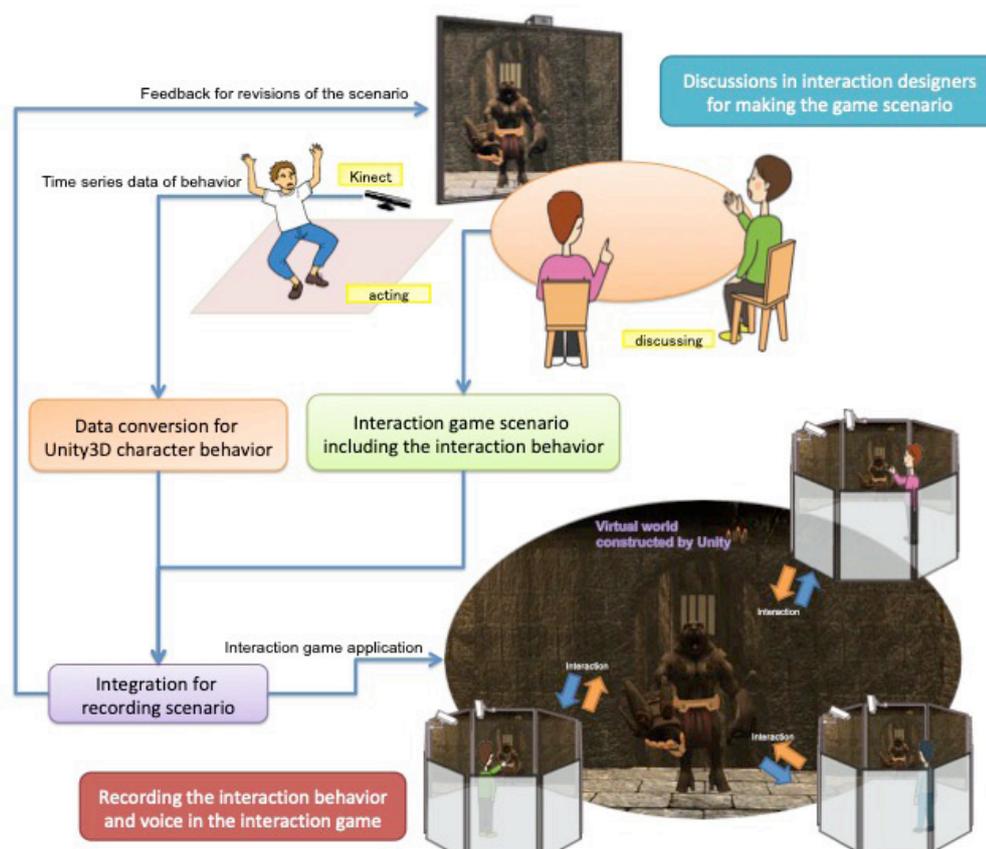


Figure 1. The immersive environment setting used in the current experiment.

3 Perceptual experiment in the audio modality

3.1 Paradigm

82 native French listeners (48F/34M, Mean age = 22.39 years) and 97 native Japanese listeners (67F/30M, Mean age = 20.42 years) were recruited in both countries. The stimuli were displayed 3 times each in audio alone condition and presented in a randomized order (54 laughs (27 spontaneous / 27 social) x 3 (repetitions) = 162 stimuli).

Before the test, subjects were informed about the definition of each type of laughter and the procedure of the experiment (Figure 2). The test was conducted individually using a GUI-based interface developed under the “OpenSesame” software (MATHÔT et al., 2012). The total duration of the session took about 25 minutes. The subjects were required to listen to each stimulus at least once (but could listen to the stimulus a second time). Then, they had to select one answer among the three possible ones: “spontaneous”, “social”, “I don’t know” (Figure 3). Definitions of the type of laughter provided in the instruction were:

- Spontaneous: it seems to you that the person is laughing in a spontaneous manner to an external event (e.g. a funny clip)
- Social: it seems to you that the person is laughing to maintain the communication with the other (e.g. embarrassed laughter, polite laughter, cynical laughter...)

Instructions

Nous vous remercions infiniment pour votre participation à notre expérience sur les rires.

Il vous est demandé d'écouter et d'identifier une série de rires produits par des locuteurs hommes et femmes.

L'expérience se déroule de la manière suivante:

Pour chaque rire:

1) Vous allez entendre la séquence une première fois

2) Vous devez choisir UNE réponse parmi les trois choix possibles:

- "Spontané": il vous semble que la personne rit de manière spontanée à un événement extérieur (Ex: une vidéo amusante)
- "Social": il vous semble que la personne rit pour maintenir le dialogue avec son (ses) interlocuteur(s). (Ex: rire embarrassé, rire poli, cynisme...)
- "Je ne sais pas"

3) Dans le cas où:

- vous avez sélectionné "Spontané" ou "Social". Veuillez indiquer à quel point vous êtes certain(e) que ce rire correspond à votre choix en cliquant sur une échelle de 1 (pas certain(e)) à 7 (tout à fait certain(e))
- vous avez sélectionné "Je ne sais pas", le rire suivant est lancé

ATTENTION:

Vous avez la possibilité de réécouter le son 1 SEULE FOIS SUPPLÉMENTAIRE. Une fois que vous avez appuyé sur "Écoute" pour écouter une seconde fois le stimulus, vous devrez faire votre choix pour passer au rire suivant

Veuillez suivre votre première intuition lors de votre choix.

Maintenant, vous allez commencer une phase d'entraînement afin de vous habituer à l'interface.
Vous passerez ensuite à la phase expérimentale.

Appuyez sur "Ok" pour commencer.
(le premier rire va se lancer directement)

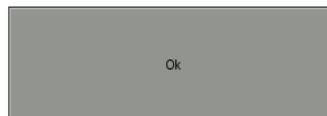


Figure 2. Part of the instructions given to the French listeners.
The interface was translated in Japanese for the Japanese listeners.

Quel est ce rire?

spontané

social

jenesaispas

Figure 3. Interface displaying the multiple choice.
From top to bottom: "spontaneous", "social", "I don't know"

3.2 Results

First of all, a χ^2 test was computed on all the responses for both groups of listeners (French and Japanese) to check whether their distributions (Social, Spontaneous or Unknown) are independent. According to the result, a significant difference for the distribution of answers was observed ($\chi^2 = 9396.5$, $df: 2$, $p < .001$).

According to Table 1 and Table 2 (stimuli are in rows and the responses given by the subjects are in columns), the two types of laughter are well recognized: French subjects identified correctly 69.24% of spontaneous laughs and 69.41% of social laughs; Japanese listeners recognized 72.25% of spontaneous laughs and 69.02% of social laughs. This result confirms that the listeners of both groups are able to recognize two types of laughter without visual indices or context.

Table 1. Results for the perceptual test for 82 French listeners.

Raw results are presented with their frequency for each row.

	Spontaneous	Social	Unknown
Spontaneous	4599 (69.24%)	1683 (25.34%)	360 (5.42%)
Social	909 (13.69%)	4444 (69.41%)	1289 (19.41%)
Total result	5508 (41.46%)	6127 (46.12%)	1649 (12.41%)

Table 2. Results for the perceptual test for 97 Japanese listeners.

Raw results are presented with their frequency for each row.

	Spontaneous	Social	Unknown
Spontaneous	5677 (72.25%)	1908 (24.28%)	272 (3.46%)
Social	1219 (15.51%)	5423 (69.02%)	1215 (15.46%)
Total result	6896 (43.88%)	7331 (46.65%)	1487 (9.46%)

3.3 Distribution of the perceptual values

In order to observe the perceptual distance of all responses based on the listeners' classification (spontaneous, social, I don't know) for the 54 stimuli, we computed a Correspondence Analysis (CA) for both Japanese and French groups using FactoMineR package (R project) under R software. According to the CA, the perceptual behavior for 26 stimuli in the French group and 20 stimuli in the Japanese one, listeners showed an important contribution (i.e. above the expected average contribution) on the first two dimensions which explain 99.8% of variances for French group and 100% of variances for Japanese group. The blue dots on the figures represent the distribution of the perceptual behavior and the three red triangles represent the concept subjects have of the types of laughter. Therefore, the closer the blue dots are to the red triangle, the more it is perceived as this type of laughter.

Similarly to the global observation, this statistical analysis shows the perceptual patterns in French and Japanese groups. In Figure 4 and Figure 5, the perceptual cluster of "sponta-

neous” laughter is located in mid left side, while the cluster for “social” laughter is located on the right bottom corner. It indicates that French and Japanese groups are able to discriminate clearly the two types of laughter showing some confusion between “social” and “unknown”.

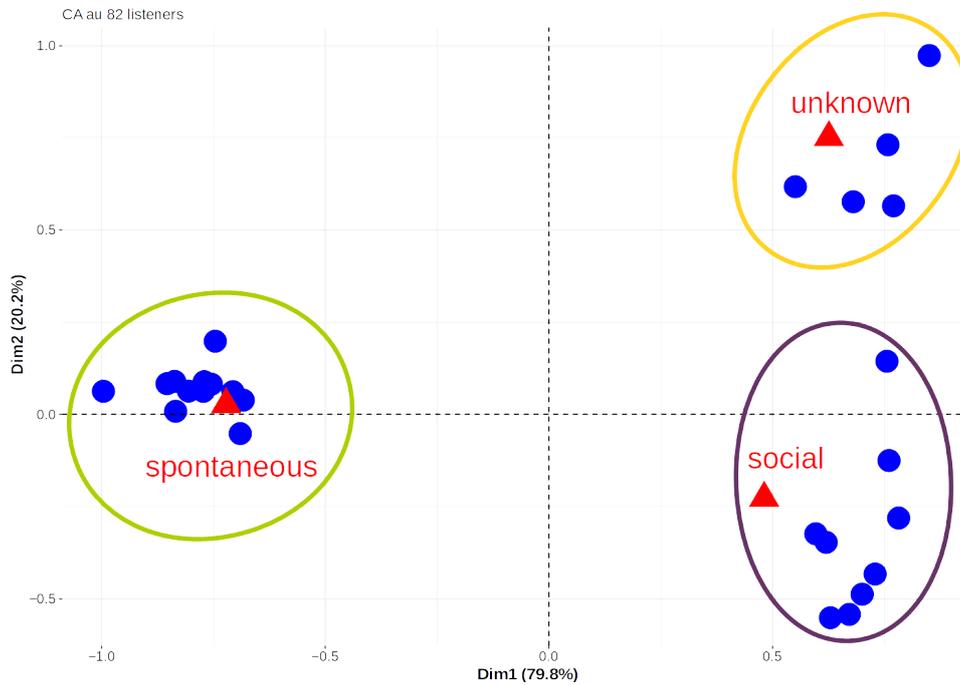


Figure 4. Distribution of the perceptual behavior of the French listeners for 26 stimuli - audio alone condition

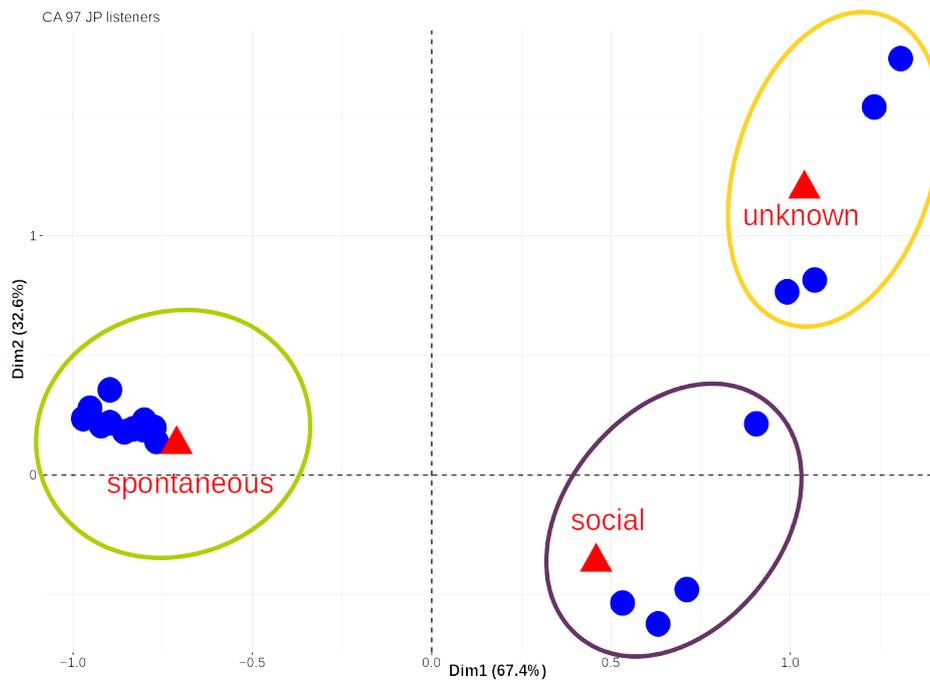


Figure 5. Distribution of the perceptual behavior of the Japanese listeners for 20 stimuli - audio alone condition

4 Perceptual experiment in the audiovisual modality

4.1 Paradigm

98 French native listeners (65F/33M, Mean age = 20.64 years) and 87 Japanese native listeners (65F/22M, Mean age = 19.74 years) were recruited in both countries. The stimuli were displayed 3 times each during the audiovisual condition in a randomized order (54 laughs (27 spontaneous / 27 social) x 3 (repetitions) = 162 stimuli). The stimuli are the same as those used for the audio alone condition. The instructions were also identical to the audio alone condition.

4.2 Results

A chi² test was computed on all the responses for both groups of listeners (French and Japanese) to evaluate whether the distributions of listeners' responses (Social, Spontaneous or Unknown) are independent. According to the result, a significant difference in the distribution of answers was observed between the two groups (chi² = 4225.1, df: 2, p < 0.001).

According to Tables 3 and Table 4, the two types of laughter are fairly recognized: French subjects correctly identified 61.64% of spontaneous laughs and 70.55% of social laughs; Japanese listeners recognized 69.39% of spontaneous laughs and 73.43% of social laughs. These results confirm that the listeners of both groups were able to recognize the types of laughter with the audiovisual as well as with the audio stimuli. Unlike the results on audio stimuli, the "social" laughs seem to be better discriminated than the "spontaneous" one for both groups. This could suggest that the subjects perceived more "social" information from visual cues.

Table 3. Results for the perceptual test for 98 French listeners.

Raw results are presented with their frequency for each row

	Spontaneous	Social	Unknown
Spontaneous	4893 (61.64%)	1715 (34.20%)	330 (4.15%)
Social	1672 (21.06%)	5601 (70.55%)	665 (8.37%)
Total result	6565 (41.35%)	8316 (52.38%)	995 (6.26%)

Table 4. Results for the perceptual test for 87 Japanese listeners.

Raw results are presented with their frequency for each row

	Spontaneous	Social	Unknown
Spontaneous	2245 (69.39%)	905 (27.94%)	86 (2.65%)
Social	663 (20.52%)	2372 (73.43%)	195 (6.03%)
Total result	2908 (44.98%)	3276 (50.67%)	281 (4.34%)

4.3 Distribution of the perceptual value

As for the responses obtained on audio stimuli, we performed the same statistical analysis on answers from the audiovisual condition. According to the CA, the perceptual behavior for 21 stimuli in the French group and 20 stimuli in the Japanese one among 54 stimuli showed an important contribution (i.e. above the expected average contribution) on the first two dimensions (100% of explained variance for both groups).

This CA analysis shows that both French and Japanese listeners have a similar perceptual behavior. In Figures 6 and 7, the perceptual groups of “spontaneous” and “social” laughs are opposed to each other on the 1st dimension. It indicates that both listeners groups are able to discriminate the two types of laughter. However, they show confusions between “social” and “unknown” for the same stimuli.

These results from audiovisual stimuli follow mostly the same perceptual pattern as the audio only condition. Therefore, it confirms that the visual information already provides an important amount of cognitive information in order to be able to discriminate two types of laughter.

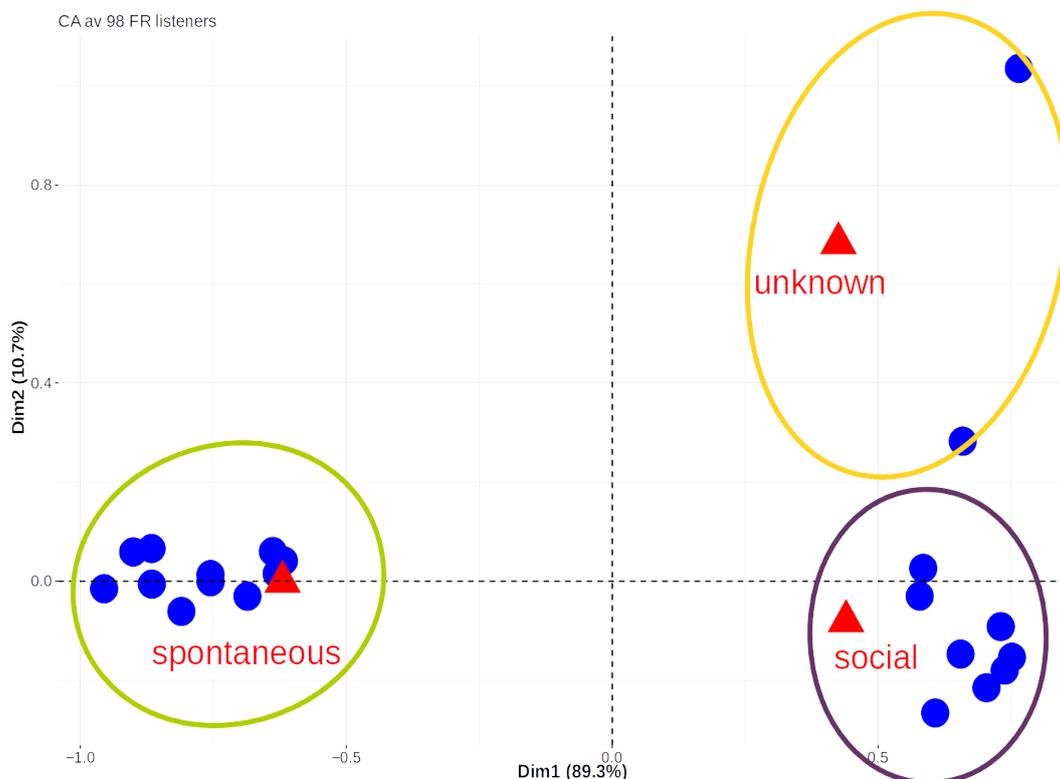


Figure 6. Distribution of the perceptual behavior of the French listeners for 21 stimuli - audio visual condition

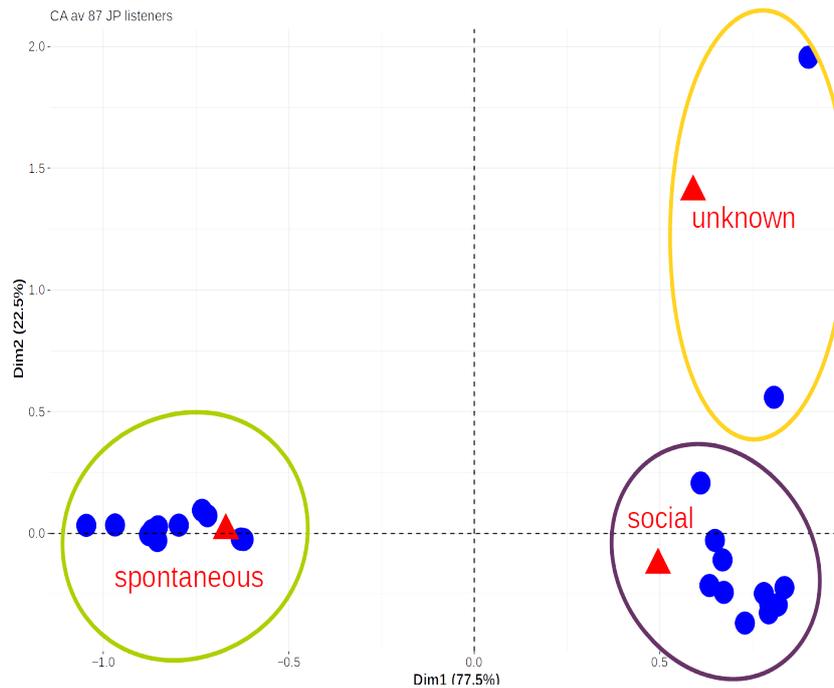


Figure 7. Distribution of the perceptual behavior of the Japanese listeners for 24 stimuli - audio visual condition

5 Acoustic analysis

Since our initial categorization of two types of laughter (mirthful spontaneous laughter and social one) is validated by both perceptual experiments (audio alone and audiovisual stimuli), we measure in the next section several acoustic features that were reported to predict affective ratings and categorization for laughter as well as for more general affective voice analysis in previous research (LAVAN et al., 2016).

5.1 Features extraction

For the acoustic analysis, the fundamental frequency (F0) and global intensity are computed every 10 ms. They are extracted using a customized version of the Snack toolkit (SJOLANDER, 2000). Most analyses are carried out on the voiced parts of the laughter as detected by the F0 extraction algorithm, thus ignoring voiceless segments.

We extracted a set of 13 features for each laugh in four main categories: F0 values for assessing the variability of the F0 (we expect, for instance, to have higher frequencies as well as more variability for spontaneous laughs), Intensity values - where higher levels and variability are also expected for spontaneous laughs, Duration values - social laughs are expected to be shorter and less voiced, Harmonics-to-noise ratios which were not explored in previous laughter studies but are expected to measure to some extent the breathiness level. The details of the features are:

- The mean value of F0 (F0 mean) extracted on voiced parts of the laughs (Hz)
- The standard deviation of F0 values (F0 sd) on a laughter excerpt (voiced parts)
- The approximated F0 slope (voiced parts only)
- The F0 minimum (F0 minimum)
- The F0 maximum (F0 maximum)
- The F0 range (F0 range)
- The mean of intensity (NRJ mean) values (dB)
- The standard deviation of intensity values (NRJ sd) during a laughter
- The intensity minimum (NRJ minimum)
- The intensity maximum (NRJ maximum)
- The intensity range (NRJ range)
- The total duration of a manually segmented laughter
- The duration of all the voiced parts of a laughter
- The percentage of voiced segments during a laughter
- The number of voiced segments
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 0.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 1.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 2.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 3.5 kHz

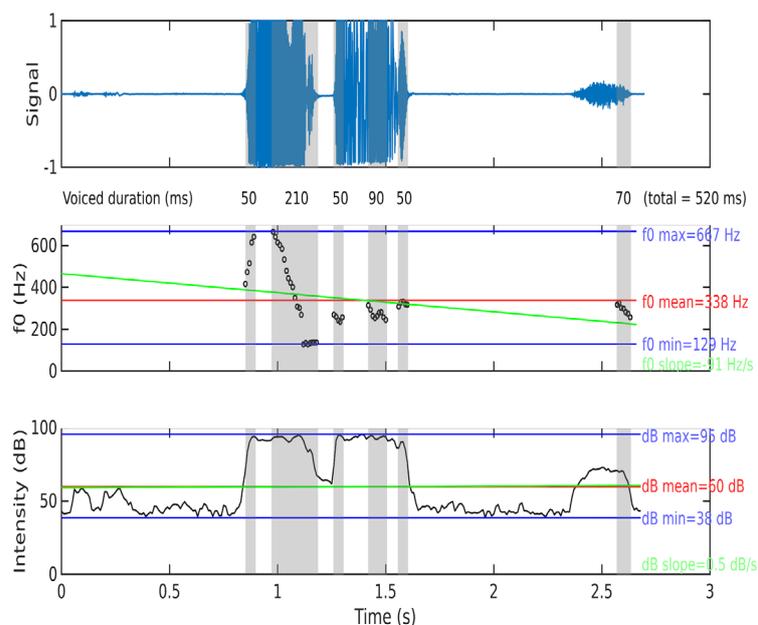


Figure 8. Extraction of acoustic features on a spontaneous laughter excerpt

5.2 Multiple Factor Analysis

To explore the global correlation between the acoustic features of F0 (mean, slope, standard deviation (sd)), intensity (mean, slope, standard deviation), total duration and voiced segment duration and the perceptual values (responses provided by the subjects) of both French and Japanese groups in audio alone condition, a Multiple Factor Analysis (MFA) was carried out. Before computing the MFA, all acoustic and perceptual values were converted into z-scores by setting the average value as reference value for each parameter. Figure 9 illustrates arrows which represent vectors of acoustic and perceptual variables. According to the MFA analysis, arrows for acoustic and perceptual variables having similar directions are considered as positively correlated variables, and are grouped together, whereas negative ones are positioned on opposite sides. This analysis shows that the distribution of the responses for French as well as for Japanese listeners are correlated with F0 features (mean and standard deviation) and the total duration of the laughter segments and of the voiced segments. However, the relative intensity (mean, slope, standard deviation) and F0 slope are less correlated with the perceptual responses of the two groups (Figure 9). Significant differences are found between spontaneous and social laughter for F0 sd ($t(52)=5.669$, $p=0.05$), for duration mean ($t(52)=2.696$, $p=0.05$) and voiced segment duration mean ($t(52)=2.595$, $p=0.05$) between spontaneous and social volitional laughs (Table 5). The variations of F0 values are higher, total duration and voiced segment duration are longer for spontaneous laughs than for social ones.

Table 5. Mean F0, F0 sd, total duration mean and voiced duration mean for the spontaneous and the social laughs

	Spontaneous	Social	t-test
F0 mean (Hz)	203.59	160.80	ns
F0 sd	54.75	25.69	2.696*
Total duration mean (s)	1.81	0.65	5.669**
Voiced duration mean (s)	0.25	0.13	2.595**

* $p<.05$, ** $p<.01$

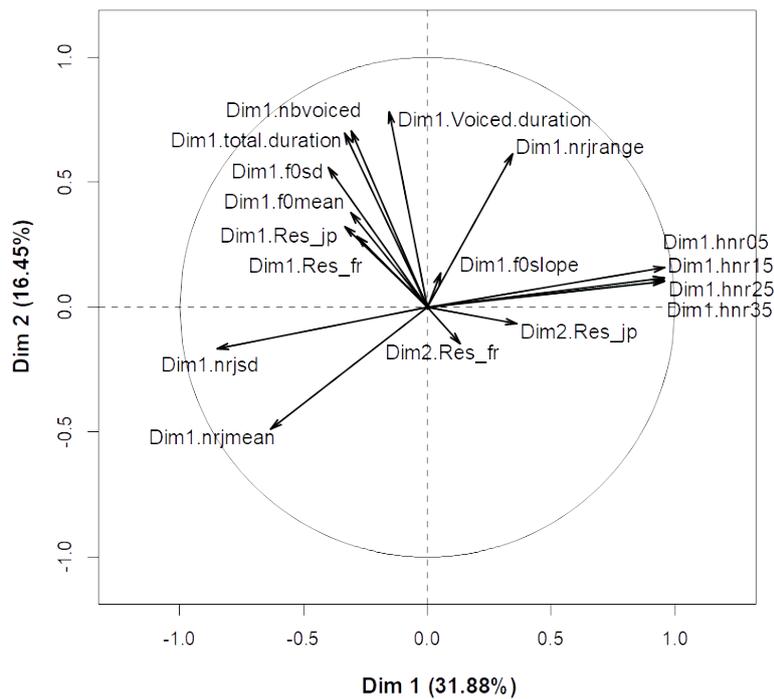


Figure 9. Correlation between acoustic and perceptual values of audio stimuli described by Multiple Factor Analysis. Vectors showing the closed direction represent positive correlation whereas negative ones are positioned on opposite sides.

5.3 Principal Component Analysis

The previous MFA analysis showed only the global correlation between all responses and all acoustic parameters. Therefore, a Principal Component Analysis (PCA) was applied to all acoustic parameters for the two types of stimuli that were categorized by all listeners (French and Japanese groups). We first analyzed all types of laughs for the intensity (mean, slope, standard deviation) and total duration (Figure 10). According to this analysis, 1st dimension (PC1) and 2nd dimension (PC2) explain 82.4% of variances. The result shows that correlations are found between the intensity standard deviation and the intensity mean vectors.

A second PCA was applied on the voiced laughs only (6 completely unvoiced laughs were removed from the set) in order to add the acoustic features related to voicing to the analysis: F0 mean, F0 slope, F0 sd, voicing duration, number of voiced segments (Figure 11). With regard to this graph, the 1st (PC1) and 2nd dimensions (PC2) explain 60.8% of variance. The result shows that the voiced duration and the number of voiced segments are correlated. F0 sd and total duration are closely correlated. Then, F0 mean and intensity slope are correlated as well. According to the distribution of the type of laughs related

to the direction of each vector on the component 1, it was found that the acoustic features concerning the voiced segment duration, the number of voiced segments, the total duration and the F0 sd help differentiate spontaneous and social laughs.

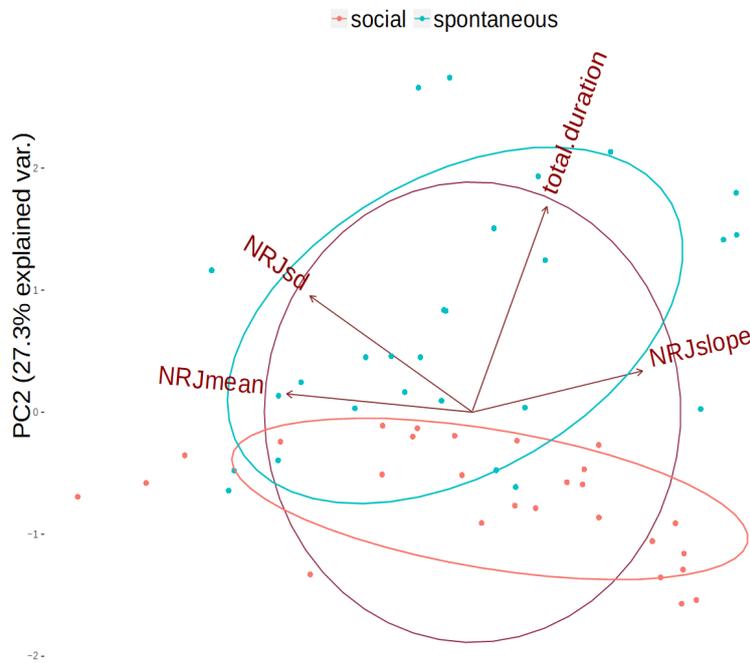


Figure 10. Correlation between the acoustic values (intensity and total duration) and the 54 laughs

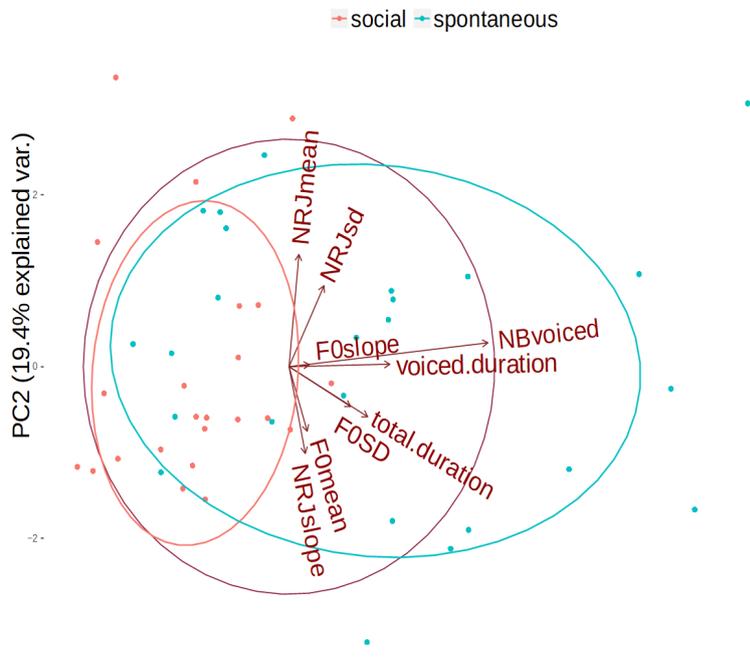


Figure 11. Correlation between the acoustic values (F0, intensity, total duration, voicing duration) and the 48 laughs

6 Subcategories of social laughter

6.1 Method

After the previous results which confirmed the initial hypothesis: human being are able to discriminate between social (voluntary) laughs and spontaneous (involuntary) ones using only audio information, we focused on the perceptual subcategorization of social laughs relying on the various psychological dimensions (LIBERMAN et al., 2007; TIAN et al., 2016; SOKOLOV; BOUCSEIN, 2000), and which may be different among languages/cultures.

Following some representative works in the field of social psychology and linguistics, three social factors: 1) speaker's physical state, 2) speaker's involvement, 3) psychological distance between speaker and partner are set up in order to create the social context in which these various social laughs occur. Each factor is evaluated on using 7 points scales on 4 dimensions as presented below:

- Speaker's physical state: calm (1) – excited (7) (dimension: arousal)
- Speaker's involvement: cooperative (1) – non-cooperative (7) (dimension: cooperation)
- Psychological distance: 2 sub-factors:
 - Close (1) – distant (7) (dimension: proximity)
 - Superior (1) – inferior (7) (dimension: hierarchy)

6.2 Experimental procedure

6.2.1 Subjects

A number of 50 French subjects (mean age = 25.9, SD= 5) and 12 Japanese subjects (mean age = 37.0, SD = 6.06) participated in the judgment task.

6.2.2 Corpus instruction

Among 27 social laughs, which were used in the previous experiment, and 14 social laughs which were best recognized by native listeners were selected as stimuli for the current perceptual experiment. An experimental interface was developed under Google form in audio alone condition. Each participant was presented with all 14 laughs in randomized order for each of 4 dimensions separately. For each factor, they were required to listen to the stimulus and rate the sound on each scale.

6.3 Results

First of all, a Kendall's coefficient of concordance was computed on the responses in order to assess the listeners' concordance of their rating each factor. The results (Table 6, Table 7) show that no general tendency was identified for both groups of listeners on the evaluation of the 14 stimuli. It indicates that the concordance of the judgments for all 14 stimuli was low, and it seems that a variety of contexts, which subjects evoked for all 14 stimuli was important for both groups.

Table 6. Results of the Kendall's coefficient of concordance for 14 social laughs rated by the 50 French listeners on the 4 dimensions

Arousal: $W=0.261114$, $\chi^2 = 169.724$ (df:13, $p.<.01$) ***
Proximity: $W=0.230408$, $\chi^2 = 149.765$ (df:13, $p.<.01$) ***
Cooperation: $W=0.266099$, $\chi^2 = 172.964$ (df:13, $p.<.01$) ***
Hierarchy: $W=0.0832044$, $\chi^2 = 54.0828$ (df:13, $p.<.01$) ***

Table 7. Results of the Kendall's coefficient of concordance for 14 social laughs rated by the 12 Japanese listeners on the 4 dimensions

Arousal: $W=0.407192$, $\chi^2 = 63.5220$ (df:13, $p.<.01$) ***
Proximity: $W=0.348421$, $\chi^2 = 54.3536$ (df:13, $p.<.01$) ***
Cooperation: $W=0.330499$, $\chi^2 = 51.5579$ (df:13, $p.<.01$) ***
Hierarchy: $W=0.157175$, $\chi^2 = 24.5193$ (df:13, $p.<.05$) *

6.4 Descriptive analysis

In this section, all responses for each dimension are described and we look for a context which corresponds to the major social laughs for both groups.

6.4.1 Arousal

Concerning the results for the French subjects for the dimension of arousal, answers ranging from 1 to 3 are considered as "calm". Answer 4 is neither calm nor excited. Answers 5, 6, 7 are interpreted as "excited". The results show that 11 laughs over 14 were judged as "calm" (the sum of the answers 1, 2, 3 constitutes more than 50% of the overall given answers). The majority of the answers are distributed among answers 2 and 3. However, 3 laughs (stimulus 3, stimulus 10 and stimulus 11) were judged as "excited". These results show that French listeners have a tendency to consider the selected social laughs as being "calm".

Japanese subjects share the same tendency for 10 laughs over 14. The four other laughs were considered as “excited” (stimulus 3, stimulus 4, stimulus 8 and stimulus 11). Among those four laughs, French and Japanese subjects share the same perceptual behavior for stimulus 3 and 11 but stimulus 4 and 8 were perceived differently.

Those results show that both groups of subjects have the tendency to globally judge social laughs as “calm”.

6.4.2 Proximity

In the dimension of the social proximity, answers 1, 2, 3 represent socially “close” between the speaker and their partner. The answer 4 is neither “close” nor “distant”. Answers 5, 6, 7 represents socially “distant” between the speaker and their partner. The results for the French listeners showed that seven laughs over 14 were judged as socially “close” between speakers. A number of 3 laughs were neither “close” nor “distant” and five laughs (stimulus 9, 12, 13 and 14) were considered socially “distant”. These results suggest that French listeners have a tendency to perceive the social relationship between speakers as “close”.

However, a majority of Japanese listeners judged seven laughs socially “distant” between speakers. Two laughs were considered neither “close” nor “distant” and five laughs were considered as socially “close”. It is important to note that even if Japanese and French subjects perceive differently the dimension of proximity, they nevertheless share the same perceptual behavior for the four stimuli considered as “distant”.

6.4.3 Cooperation

For the results about cooperation, answers 1, 2, 3 are considered as “cooperative”. Answer 4 is neither “cooperative” nor “not cooperative”. Then answers 5, 6, 7 are considered as “not cooperative”. The results for the French listeners show that seven laughs over 14 are interpreted as “cooperative”. A number of three stimuli were judged as neither “cooperative” or “not cooperative” and four stimuli were considered as “not cooperative”. This result shows the tendency of the French subjects to consider the social laughs as being “cooperative”.

Similar to the French subjects, Japanese subjects judged the speaker’s attitude which appeared in the majority of social laughs (6 over 14 laughs) as being “cooperative”, three of them were interpreted as neither “cooperative” or “not cooperative”, and five laughs as “not cooperative”. Moreover, both of the groups share the same perceptual behavior by interpreting the same set of stimuli as being “not cooperative”.

6.4.4 Hierarchy

On the hierarchy dimension, answers 1, 2, 3 are considered as socially “superior” to the conversation partner. Answer 4 is interpreted as neither “superior” nor “inferior”. Answers 5, 6, 7 are interpreted as socially “inferior” to the partner. For the French group, the majority of the stimuli were considered as neither “inferior” nor “superior” with eight laughs over 14. Then, the distribution of the answers for the 6 other stimuli is equally distributed between “superior” and “inferior”. This result shows that French subjects recognized the social hierarchy between the speaker and the interlocutor as neither “superior” nor “inferior”.

As for the French group, Japanese subjects considered the social laughs as neither “superior” nor “inferior” for seven laughs over 14. More stimuli were considered as “superior” for the Japanese subjects with a number of five laughs and only two stimuli were interpreted as “inferior”. It indicates that both Japanese and French subjects share the same tendency: they globally judged the social relation between speakers as neither “superior” nor “inferior”.

7 Conclusion

The current paper investigates whether human beings can perceptually discriminate between social volitional laughter and spontaneous involuntary laughter from a corpus of spontaneous laughs recorded in a virtual immersive environment using only auditory laughs or audiovisual laughs without any context or any foreign language skill. According to the perceptual discrimination experiment with only auditory laughs, two groups of subjects (i.e. native French and Japanese subjects) who belong to quite different cultures/languages are able to discriminate these two types of laughter indicated by more than twice the chance level of recognition rate, without visual information nor any context. In the second perceptual experiment with audiovisual laughs, both groups discriminated well showing a high recognition rate as much as in the previous audio alone condition. These results confirm the existence of two types of laughter on the voluntary-involuntary control dimension as mentioned in previous research (SCOTT et al., 2014; LAVAN et al., 2016).

Thirteen acoustic features including F0, harmonic to noise ratio, intensity and duration for each type of laughter are also investigated. It is important to mention that the acoustic variation of laughs due to the speaker’s native language (i.e. French and Japanese)

was not verified for this current experiment. Indeed, it was difficult to compare such a cultural factor because of the lack of French subjects' samples and of the importance of the intrinsic variation among all laughs. Therefore, we focused on the global analysis excluding speech variation due to the cultural factor. First of all, Multiple factor Analysis was conducted to explore the global correlation between the acoustic characteristics and the participants' perceptual behavior. Results show that the perceptual behaviors of both French and Japanese groups are correlated with F0 features (mean and standard deviation), the total duration and the voiced segment duration. After this global result, we further investigated the important acoustic factors associated to each type of laughter (spontaneous or social). The results show that the total duration helps to differentiate spontaneous laughs from the social ones. Moreover, we found that the voiced duration, the number of voiced segments and the F0 standard deviation also contribute to the differentiation between spontaneous and social laughs.

We also investigated subcategories of social laughs on a subset of excerpts validated in the previous perceptual experiment. More concretely, we implemented a third perceptual experiment with 14 highly rated social laughs for Japanese and French groups. According to the psycho-dimensional analysis, we are able to describe a context for French group: the speaker knows the conversation partner. Both speaker and his partner are students. The speaker is willing to go on the conversation (and social interaction). In this context, speaker laughed in calm. Japanese subjects evoked a similar context except for the nature of psychological distance: the speaker does not know well the conversation partner. Both speaker and his partner are students. Speaker is willing to go on the conversation (and social interaction). In this context, speaker laughed in calm.

Acknowledgments

This study has been carried out with financial support from the French State in the frame for the "Investments for the future" program IdEx Bordeaux (ANR-10-IDEX-03-02), the Bordeaux PEPS IDEX/CNRS project "Virtual Laughter", the Bordeaux International Support in collaboration with Tsukuba University "Cultural difference of Social Laughter". We also wish to thank all the speakers and listeners from Kyoto University, Ritsumeikan University and Bordeaux Montaigne University for their kind participation for their contribution.

References

- ANIKIN, A.; LIMA, C.F. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology*, p. 1-21, 2017.
- BACHOROWSKI, J.A.; SMOSKI, M. J.; OWREN, M. J. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, v. 110, n. 3, p. 1581-1597, 2001.
- BERK, R. A. The active ingredients in humor: Psychophysiological benefits and risks for older adults. *Educational Gerontology*, v. 27, no 3-4, p. 323-339, 2001.
- BOERSMA, P.; DAVID, W. *Praat, a system for doing phonetics by computer* [Computer program], Version 5.3.51, 2013.
- BUTLER, B. *Laughter The Best Medicine?* Oregon Institute of Marine University of Oregon Biology, ed. Spring, v. 11, n. 1, p. 1, 2005.
- CAMPBELL, N.; KASHIOKA, H.; OHARA, R. No laughing matter. In: Interspeech 2005, Lisbon, Portugal, 2005, p. 465-468.
- ERICKSON, D.; MENEZES, C.; SAKAKIBARA, K. Are you laughing, smiling or crying? In: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, Sapporo, Japan, 2009. p. 529-537.
- FRÓIS, L., S.J. *Traité sur les contradictions et différences de mœurs écrit par le R. P. Luís Fróis au Japon, l'an 1585, traduit du portugais, Paris, ed. Chandeigne, 1998.*
- ISHI, C.; HATANO, H.; ISHIGURO, H. Audiovisual analysis of relations between laughter types and laughter motions. In : Proc. of the 8th international conference on Speech Prosody (Speech Prosody 2016). 2016. p. 806-810.
- JACYKIEWICZ, T.; RINGEVAL, F. *Automatic recognition of laughter using verbal and non-verbal acoustic features*. 2014. Ph.D. dissertation - Department of Informatics, University of Fribourg, Switzerland, 2014.
- JÜRGENS, R.; HAMMERSCHMIDT, K.; FISCHER, J. Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in psychology*, v. 2, p. 180, 2011.
- KIPPER, S.; TODT, D. Variation of sound parameters affects the evaluation of human laughter. *Behavior*, v. 138, n. 9, p. 1161-1178, 2001.
- LAVAN, N.; SCOTT, S.K.; MCGETTIGAN, C. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, v. 40, n. 2, p. 133-149, 2016.
- LIBERMAN, N.; TROPE, Y.; STEPHAN, E. Psychological distance. In: *Social Psychology Handbook of basic principles*. Guilford Press, v. 2, 2007. p. 353-381.
- MATHÔT, S.; SCHREIJ, D.; THEEUWES, J. Opensesame: An opensource, graphical experiment builder for the social sciences. *Behavior research methods*, v. 44, n. 2, p. 314-324, 2012.

OWREN, M.; BACHOROWSKI, J. Acoustic assessment of vocal expression of emotion. In: *Handbook of emotion elicitation and assessment*, Oxford University, 2007. p. 239-266.

R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna, Austria, 2017. Available at: <https://www.R-project.org>.

PROVINE, R. R. Laughter. *American Scientist*, v. 84, p. 38-45. 1996.

PROVINE, R. R. *Laughter: a scientific investigation*. New York: Penguin, 2001.

SCOTT, S. K. et al. The social life of laughter. *Trends in cognitive sciences*, v. 18, n. 12, p. 618-620, 2014.

SHOCHI, T.; RILLIARD, A. La prononciation des apprenants de fle et la multimodalité expressive. *La prononciation du français dans le monde*, ed. CLE International, 2016. p. 257-263.

SJOLANDER, K. *The Snack Sound Toolkit* [Computer program], 2000.

TANAKA, H.; CAMPBELL, N. Acoustic features of four types of laughter in natural conversational speech. Proc. 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 2011. p. 1858-1961.

TROUVAIN, J. Segmenting phonetic units in laughter. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2003. p. 2793-2796.

SOKOLOV, E.N.; BOUCSEIN, W. A. Psychophysiological model of emotion space. *Integrative Physiological and Behavioral Science*, v.35, n. 2, p. 81-119, 2000.

TIAN, Y.; MAZZOCCONI, C.; GINZBURG, J. When do we laugh? In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2016. p. 360-369.

VETTIN, J.; TODT, D. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behavior*, v. 142, n. 2, p. 217-240, 2005.

WICHMANN, A. The attitudinal effects of prosody, and how they relate to emotion. In: International Speech Communication Association Tutorial and Research Workshop on Speech and Emotion, 2000.

WILD, B. et al. Neural correlates of laughter and humour. *Brain*, v. 126, n. 10, p. 2121-2138, 2003.



Submitted on: 14/03/2018

Accepted on: 03/06/2018

O PAPEL DA FONOLOGIA DO PORTUGUÊS BRASILEIRO NA TRADUÇÃO DE TEXTOS MUSICAIS

THE ROLE OF THE BRAZILIAN PORTUGUESE PHONOLOGY
IN THE TRANSLATION OF SINGABLE LYRICS

Graziela Pigatto Bohn | [Lattes](#) | grazielabohn@gmail.com
Universidade Católica de Santos

Roseane Silva da Cruz | [Lattes](#) | roseanescruz@gmail.com
Universidade Católica de Santos

Resumo: Este estudo investiga o processo tradutológico de músicas para o canto, verificando quais estratégias o versionista lança mão para que a tradução seja tanto logocêntrica como melocêntrica quando comparada à canção original. Para esse fim, fizemos uma comparação entre a notação musical da versão original da música *Seasons of Love*, composta por Jonathan Larson para o teatro musical *Rent* em 1994, e sua versão para o português brasileiro, realizada pela versionista Mariana Elisabetsky em 2016. Com base nos aspectos da fonologia do português brasileiro, em especial ritmo, acento e os processos de juntura segmental (COLLISCHONN, 2001), nos conceitos de acento métrico e ritmo sob a perspectiva da música (BENNETT, 1987; LACERDA, 1967), e nos processos tradutológicos de músicas para o canto (FRAZON, 2008), nossa análise mostra que não é possível a realização de uma tradução literal de músicas para o canto, o que leva o versionista a criar uma adaptação para que a letra traduzida possa ser acompanhada junto à notação musical de sua letra original e, ao mesmo tempo, não cause estranhamento ao ouvinte.

Palavras-chave: Versionista; Tradução de música para o canto; Português brasileiro; Notação musical; Teatro musical.

Abstract: This paper aims to investigate the translation process of singable lyrics, verifying which strategies the translator uses so the translated song remains both logocentric and melocentric when compared to the original. For this purpose, we compared the original sheet music of the song “Seasons of Love” by Jonathan Larson, from the musical “Rent” in 1994, with its Brazilian Portuguese translation by Mariana Elisabetsky, in 2016. On the

basis of the aspects of Brazilian Portuguese phonology, in particular rhythm, stress and the processes of segmental junction (COLLISCHONN, 2001), in the concepts of metric accent and rhythm from the music perspective (BENNETT, 1987; LACERDA 1967), and with the translation process of singable lyrics (FRAZON, 2008), we could see that the literal translation of a song is not possible when its intention is to be sung. Therefore, the translator needs to adjust the translated song in order to fit its lyrics into the melody, so the translated song can be naturally sung along with its original sheet music.

Keywords: Translator; Singable translation; Brazilian Portuguese; Sheet music; Musical theatre.

Introdução

Nos últimos anos, o ramo cultural do teatro musical no Brasil tem se expandido e importado cada vez mais montagens musicais renomadas de outros países, em especial dos teatros da Broadway (*Broadway Theater*) ou dos teatros da *West End* de Londres, dois dos mais representativos teatros de musicais em língua inglesa. As montagens brasileiras desses musicais exigem adaptações que transmitam a mesma informação agregada à montagem original em língua estrangeira. Por esse motivo, o versionista exerce um papel fundamental no processo de tradução de peças musicais, pois, no teatro musical, a história é contada, principalmente, por meio das canções. O versionista deve ainda preocupar-se em manter o mesmo esquema de rimas, repetições e estrofes, uma vez que a música original e a tradução estão submissas à mesma notação musical.

Neste estudo, buscamos analisar de que forma aspectos fonológicos da língua de chegada podem servir como uma ferramenta para que a canção versionada corresponda melocentricamente à canção original preservando o seu conteúdo.

Na primeira seção, trataremos dos aspectos que a tradução de músicas deve apresentar para que esta seja considerada cantável, de acordo com Frazon (2008). Na segunda seção, abordaremos os pontos teóricos musicais mais relevantes para o processo tradutológico de canções. Na terceira seção, discutiremos os aspectos fonológicos do português brasileiro, geralmente usados como ferramenta no trabalho do versionista. Na quarta seção, por meio da análise da versão brasileira da canção *Seasons of Love*, traduzida como “Dias de Amor” do teatro musical *Rent*, apresentamos nossos resultados, que são seguidos pelas considerações finais.

A tradução de textos musicais para o canto

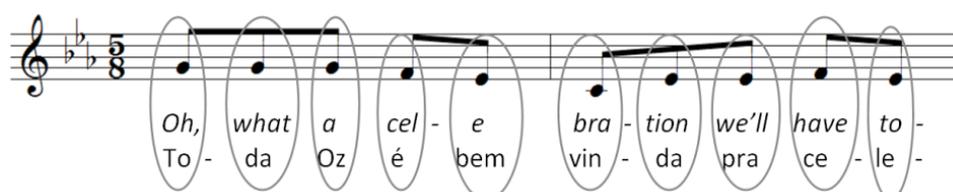
O termo texto musical refere-se à letra de canções. Todo texto musical traduzido para o canto é chamado de versão musical, e esse tipo de tradução é comissionado a profissionais versionistas, que são geralmente compositores, cantores, especialistas em óperas e dramaturgos. Uma vez que essa área requer conhecimentos teóricos musicais, são poucos os profissionais de tradução que atuam nesse campo.

Uma das tarefas do versionista é a de buscar soluções na língua de chegada para que a letra e a melodia se encaixem precisamente sem que haja a necessidade de alterar a partitura do texto musical, pois a tradução dos textos musicais está submissa à notação musical da canção original.

Segundo Frazon (2008, p. 391), a tradução de canções passa por um processo músico-textual que envolve três camadas de cantabilidade (*the three layers of singability*), cada uma das quais representando a finalidade que essa tradução deve atingir para que o texto musical seja considerado cantável.

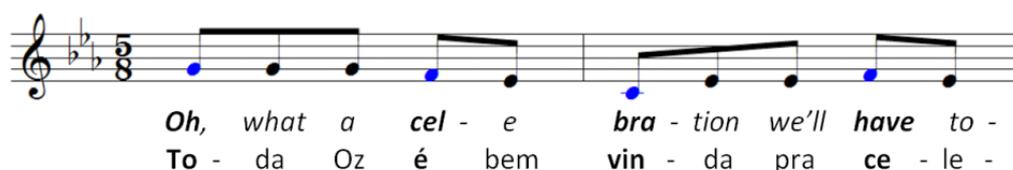
A primeira camada é onde ocorre a combinação prosódica. De acordo com Frazon (p. 393), essa é a camada de maior relevância para a cantabilidade do texto musical traduzido, pois nela se realiza a correspondência métrica da canção. É a camada que torna o texto musical cantável, ainda que, neste estágio, a letra possa não fazer sentido. Por exemplo, na partitura na Figura 1, abaixo, os símbolos que representam as notas musicais e as palavras correspondem-se silabicamente, ao mesmo tempo que as tônicas das palavras devem corresponder aos tempos fortes da música, conforme Figura 2. Dessa correspondência dá-se a combinação prosódica, sendo possível acompanhar, simultaneamente, o texto musical original e traduzido sob a mesma notação musical (FRAZON, 2008, p. 393), conforme ilustramos a seguir:

Figura 1. Correspondência silábica



Fonte: elaborada pelas autoras.

Figura 2. Correspondência tônica e tempos fortes



Fonte: elaborada pelas autoras.

Na segunda camada, temos a combinação poética, que diz respeito aos recursos estilísticos da poesia, mas que, sozinha, não torna o texto musical cantável, o que faz com que a segunda camada seja dependente da combinação prosódica presente na primeira camada. Por exemplo, no Quadro 1, a seguir, podemos perceber que a terceira coluna é logocêntrica, pois foca na tradução literal de palavras e falha ao estabelecer correspondência com os aspectos métricos da canção original.¹

Quadro 1. Letra português / inglês de “Popular” de *Wicked*

ORIGINAL	VERSÃO PARA O CANTO	TRADUÇÃO LITERAL
Popular, you're gonna be popular! I'll teach you The proper loys When you talk to boys Little ways To flirt and flounce	Popular, te ensino a ser popular Uns truques de quem testou E se aprimorou Estratégias pra flertar	Popular, você será popular! Eu vou te ensinar Os truques certos Para falar com os garotos Pequenas maneiras De flertar e se movimentar

Fonte: elaborada pelas autoras.

Na terceira camada, temos a combinação semântica, que diz respeito ao modo como as palavras relacionam-se mediante rimas, repetições e estrofes, e os efeitos de sentido que elas criam por meio da oralidade, manipulando as emoções do público. Frazon (2008, p. 392) nos diz que a combinação semântica é atingida quando o clímax melódico coincide com o clímax imposto no texto musical, dando significado à canção, isto é, letras felizes devem ser acompanhadas por uma melodia alegre, assim como às letras tristes são atribuídas melodias melancólicas; o texto musical reflete, portanto, o movimento que a música expressa.

¹ É comum que o idioma do texto musical original e do texto musical traduzido possuam divergências tanto sintáticas como fonológicas, levando o versionista a produzir uma adequação contextual na qual a informação é mantida, porém não é possível uma comparação tradutológica palavra por palavra entre os textos. A fim de justificar as escolhas tradutórias do versionista, Frazon (2008) refere-se à tradução de textos musicais cantáveis como adaptações.

O versionista deve preocupar-se, ainda, com que o texto musical apresente uma tradução persuasiva e expressiva, sendo ela o efeito das atribuições devidas dos seus valores prosódicos e poéticos. A tradução do texto musical para o canto cumpre com suas exigências quando há a combinação semântica, pois a terceira camada é o produto final da junção das duas camadas anteriores.

Elementos formais da música

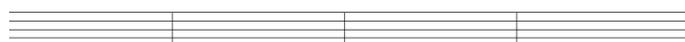
Vimos, na seção anterior, que, conforme Frazon (2008), o versionista deve cumprir algumas exigências a fim de que a tradução dos textos musicais seja cantável. A combinação prosódica é o primeiro item para que se atinja a cantabilidade, pois a partir dessa combinação é que obtemos a correspondência entre a sílaba e a notação musical, correspondência métrica entre a tônica da palavra e os tempos fortes da música; e é por meio dessa combinação, também, que o ritmo e a entonação da canção tornam-se perceptíveis.

Nesta seção, trataremos de alguns dos aspectos teóricos musicais – notação musical, acento métrico e ritmo – que o versionista deve levar em conta para que a tradução da música em questão seja apta ao canto.

Notação musical

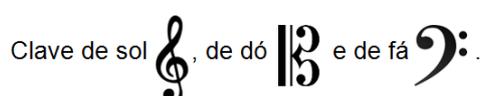
No processo de tradução de textos musicais há um sistema musical textual que o versionista deve seguir para que a tradução do texto seja apropriada ao canto (FRAZON, 2008, p. 393). Uma das etapas é a correspondência silábica, de acordo com a qual cada sílaba cantada deve coincidir com uma notação musical. Segundo Bennett (1987), a notação musical é a representação do som por meio de símbolos, sendo alguns deles os que apresentamos nas Figuras 3 – 8, abaixo:

Figura 3. Pauta ou pentagrama



Fonte: Bennett (1987, p. 11, adaptada).

Figura 4. Claves



Fonte: Bennett (1987, p. 10, adaptada).

A clave (Figura 4), dependendo de sua posição na pauta, é a que determina o nome das notas: dó, ré, mi, fá, sol, lá, si.

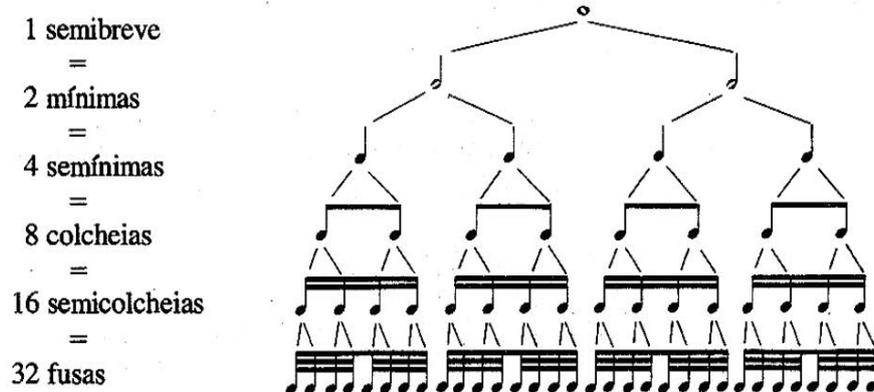
Figura 5. Notas e suas respectivas pausas



Fonte: Bennett (1987, p. 11 , adaptada).

As figuras de notas e pausas, dependendo de seu formato, têm a duração longa ou curta em relação à outra. A figura de semibreve, por exemplo, é a que possui o maior tempo de duração e abrange todas as demais:

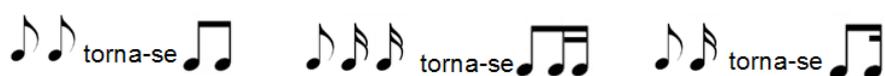
Figura 6. Valor relativo das notas



Fonte: Bennett (1987, p. 12).

Quando juntas no mesmo compasso, colcheias, semicolcheias, fusas e semifusas, podem apresentar o seguinte formato:

Figura 7. Colcheia e semicolcheia



Fonte: Bennett (1987, p. 12).

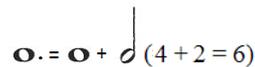
Para um som prolongado, quando as notas são da mesma altura, usa-se a *ligadura de prolongamento* (Figura 8), que faz com que a duração do som seja a soma da duração das notas ligadas; ou o *ponto de aumento* (Figura 9), que adiciona à nota metade do seu valor. Quando as notas ligadas são de alturas diferentes usa-se uma *ligadura de expressão* (Figura 10), a qual não altera o tempo de duração das notas.

Figura 8. Ligadura de prolongamento



Fonte: Bennett (1987, p. 12 , adaptada).

Figura 9. Ponto de aumento



Fonte: Bennett (1987, p. 12).

Figura 10. Ligadura de expressão



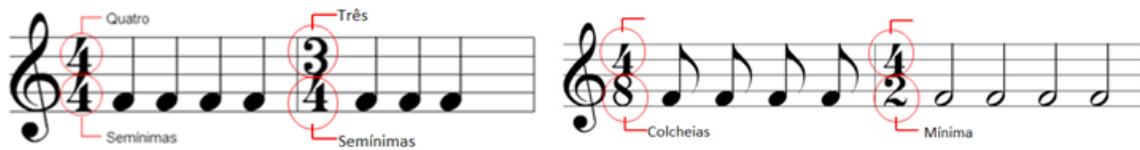
Fonte: Bennett (1987, p. 12 , adaptada).

A ligadura de prolongamento influencia os textos musicais, pois as notas ligadas, geralmente, são marcadas como uma só, ou seja, notas ligadas, preferencialmente, devem ser associadas a uma única sílaba.

Acento métrico

Na música, o acento métrico é definido a partir dos tempos fortes e fracos do compasso (LACERDA, 1987). O tipo de compasso pode ser identificado no início da partitura e é representado por uma fração chamada de *fórmula de compasso*. Nessa fração, o numerador indica a quantidade de tempos de cada compasso e o denominador indica a nota que terá a duração de 1 tempo (BENNETT, 1987, p. 13):

Figura 11. Fórmula de compasso



Fonte: Bennett (1987, p. 13 , adaptada).

Normalmente, o tempo forte (*f*) é o primeiro tempo de cada compasso, sendo os tempos restantes denominados meio-forte (*mf*) ou fraco (*p*) (LACERDA, 1967, p. 15-16), Vejamos alguns exemplos:

Figura 12. Compasso binário simples

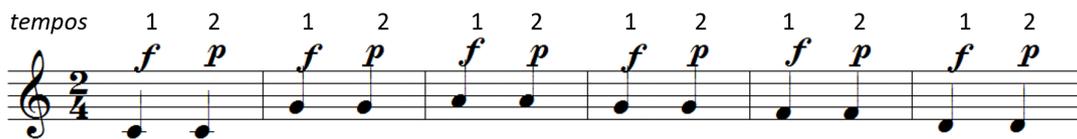
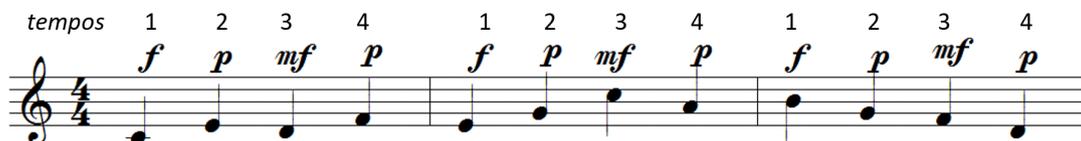


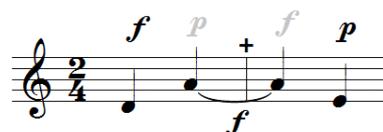
Figura 13. Compasso quaternário simples



Fonte: Lacerda (1967, p. 15-16 , adaptada).

Na composição, há recursos que podem ser usados para destacar certos pontos da música. Um desses recursos é a *síncopa*², que ocorre quando há a presença da ligadura de prolongamento (LACERDA, 1967, p. 38). O tempo fraco da música é ligado ao tempo forte seguinte, criando um deslocamento da acentuação.

Figura 14. Síncopa



² Na linguagem musical, o termo é usado no gênero feminino.

Fonte: Lacerda (1967, p. 38 , adaptada).

Assim, podemos perceber que o acento métrico é definido a partir das batidas padrões fortes ou fracas da música e segue um parâmetro que raramente pode ser alterado. Para que o texto musical traduzido seja cantável, além da correspondência silábica, os tempos forte e meio-forte da música devem, preferencialmente, coincidir com a sílaba tônica das palavras (FRAZON, 2008, p. 393).

Ritmo

Enquanto a métrica pode ser definida pelos tempos fortes e fracos da música, o ritmo está relacionado à duração; ou seja, de acordo com Med (1984), as combinações entre as diferentes possibilidades de duração do som e silêncio – *notas* e *pausas*, organizados em proporções exatas – resulta no ritmo musical, sendo essa combinação responsável pela criação e caracterização de diferentes melodias.

Para Benward & Saker (2009), ritmo e métrica operam em conjunto, sendo o ritmo um modelo de *durações* irregulares, enquanto que a métrica é formada por batimentos e medidas padrões:

Figura 15. Métrica e ritmo



Fonte: Benward e Saker (2008, p. XVI).

Vale notar a contribuição de Hasty (1997, p. 4) no que diz respeito ao ritmo e à métrica. Segundo o autor, a métrica é fixa, e o ritmo é livre, uma vez que a métrica é uma medida padrão determinada e o ritmo é o movimento musical, influenciado pelos sentimentos e emoções do compositor e pela forma de expressão e execução do intérprete.

Medeiros (2009) justifica que o movimento da música é marcado pela melodia e as durações são seus organizadores, isto é, ainda que uma música seja reproduzida em uma velocidade diferente, as proporções das durações que a caracterizam são mantidas mentalmente de modo que a música executada seja reconhecida pelo ouvinte. Por essa razão,

um dos métodos utilizados, por exemplo, é a atribuição de duas ou mais notas ligadas para sílabas longas, ou em final de frases para a criação de um ritmo melódico, uma vez que a música e a articulação das palavras cantadas são interdependentes.

Propriedades fonológicas do português brasileiro

Nesta seção, serão abordadas algumas propriedades fonológicas do português brasileiro a fim de entendermos como o versionista pode lançar mão de processos para fazer a correspondência entre letra e música. São elas: ritmo, acento e processo de junção.

Ritmo

Para Medeiros (2009), o vínculo entre o ritmo musical e o ritmo linguístico é que em ambos o ritmo é um processo mental gerado a partir das práticas de movimentos longos ou curtos que apresentem uma padronização.

No ponto de vista linguístico, o ritmo é marcado pelas batidas da fala e suas proeminências. Essas proeminências determinam o acento das palavras: uma sílaba mais proeminente é denominada sílaba tônica, e uma sílaba menos proeminente é denominada sílaba átona (COLLISCHONN, 2001). O ritmo da fala é, também, marcado pela divisão do tempo em partes iguais, chamada de isocronia. Por meio da isocronia, línguas como o inglês, o russo e o português europeu são denominadas línguas de ritmo acentual, enquanto línguas como o francês, o japonês e o espanhol, línguas de ritmo silábico (MIGLIORINI; MASSINI-CAGLIARI, 2010).

O ritmo acentual, em (1), abaixo, é definido pela duração, sendo as sílabas tônicas longas e as sílabas átonas reduzidas. Essa duração, contudo, é dependente do tamanho do constituinte: quanto maior a palavra, menor a duração de cada sílaba, quanto menor a palavra, maior a duração. Já o ritmo silábico, em (2), abaixo, é definido pela intensidade: as sílabas tônicas são fortes e as sílabas átonas fracas, tendo cada sílaba, no geral, o mesmo tempo de duração (CAGLIARI, 2012, p. 41).

(1)	Cats	chase mice	=	3 sílabas / 3 tempos
	The cats	chase mice	=	4 sílabas / 3 tempos
	The cats	chase the mice	=	5 sílabas / 3 tempos
	The cats will	chase the mice	=	6 sílabas / 3 tempos
	The cats will have	chased the mice	=	7 sílabas / 3 tempos

(2) Gatos	perseguem ratos	=	7 sílabas / 7 tempos
Os gatos	perseguem ratos	=	8 sílabas / 8 tempos
Os gatos	perseguem os ratos	=	9 sílabas / 9 tempos
Os gatos	perseguirão os ratos	=	10 sílabas / 10 tempos
Os gatos	terão perseguido os ratos	=	12 sílabas / 12 tempos

O ritmo do português brasileiro não possui, ainda, uma definição muito clara se é acentual ou silábico. Entretanto, Migliorini & Massini-Cagliari (2010) e Cagliari (2012) argumentam que, apesar de as línguas serem divididas em dois grupos rítmicos, silábico e acentual, há imprecisões; categorizar uma língua como puramente silábica ou acentual não é possível, já que o ritmo é de caráter individual, idealizado na mente do falante (CAGLIARI, 2012).

Acento

No que diz respeito à fala, o acento está relacionado à proeminência relativa das sílabas que constituem as palavras. Embora existam exceções, há indícios acerca da regularidade no que diz respeito à acentuação tônica das palavras do PB: o acento é limitado às três últimas sílabas da palavra, o que as classifica em *oxítonas* – o acento cai na última sílaba (anZOL); *paroxítonas* – o acento cai na penúltima sílaba (ciDAde); e *proparoxítonas* – o acento cai na antepenúltima sílaba (PÂNtano).

Segundo Collischonn (2001), as oxítonas apresentam ainda duas classificações: oxítonas terminadas em consoantes e oxítonas terminadas em vogais, sendo as palavras terminadas em consoante mais frequentes em receber o acento oxítono do que aquelas que terminam com vogais (2001, p. 134). Isso porque a relação não está entre a palavra e o acento, mas, sim, entre a sílaba e o acento. O fato de a sílaba terminar em consoante a faz ser pesada, atraindo o acento (2001, p. 135):

(3) SUga	→	suGAR
COLhe	→	coLHER
MUla	→	muLHER
ANjo	→	anZOL

O acento paroxítono é o mais natural para o falante nativo da língua. E, por ser a formação de paroxítonas a mais comum do PB, é normal que novas palavras introduzidas na língua respeitem esse padrão (COLLISCHONN, 2001, p. 133).

- (4) SANDwich → sanduÍche
FACEbook → faceBUki
CInema → ciNEMA

Já as proparoxítonas são as menos frequentes, por tratar-se de um grupo formado a partir de palavras originárias do Latim e do Grego. É comum, ainda, sua modificação por meio de um processo de apagamento da penúltima sílaba, reclassificando-as em paroxítonas. Por exemplo, é natural aos falantes nativos do português brasileiro falar [ˈʃi.kra] em vez de [ˈʃi.ka.ra] para ‘xícara’ e [aˈbɔ.bra] em vez de [aˈbɔ.bo.ra] para ‘abóbora’ (COLLISCHONN, 2001, p. 133). Para Collischonn, “o acento proparoxítono é marcado, no sentido de que é menos usual. É um acento especial, contrário à tendência geral de acentuar a penúltima sílaba.” (COLLISCHONN, 2001 p. 133).

Um aspecto que também influencia a marcação de acento no PB é a junção de afijos à palavra. Por exemplo, o acréscimo de um sufixo derivacional em uma palavra fará com que a posição do acento seja alterada. Por outro lado, em casos de prefixo e sufixo flexional de número em nomes, a posição do acento, geralmente, permanecerá a mesma (COLLISCHONN, 2011, p. 136):

(5) SUFIXO DERIVACIONAL	SUFIXO FLEXIONAL	PREFIXO
FAca > faCAda	FAca > FAcas	SOlo > subsolo
ÁRvore > arvoREdo	ÁRvore > ÁRvores	inteliGENte > superinteliGENte
poLÍcia > policiAL	poLÍcia > poLÍciais	moDERno > ultramoDERno
capitAL > capitaLISTa	capiTAL > capiTAIS	CARga > sobreCARga

Vale observar, nos exemplos em (5), acima, que, no caso dos sufixos derivacionais, a possibilidade de alteração do acento permite que as palavras ajustem-se à tendência de acentuação natural da língua: *ÁRvore* e *capiTAL* tornam-se proparoxítonas com o acréscimo do sufixo derivacional: *arvoREdo* e *capitaLISTa*.

Processos de juntura

Além da marcação de acento da língua, interessam-nos também os processos fonológicos que afetam as cadeias de palavras do PB, como os de juntura segmental, em especial aqueles que afetam os núcleos silábicos, resultando em ressilabação, podendo servir como ferramenta para que haja correspondência entre tempo musical e sílabas.

O primeiro processo que abordaremos aqui trata do apagamento da vogal átona /a/ em posição final de palavra, quando a vogal inicial da palavra seguinte não portar o acento

da frase (BISOL, 1996, p. 160). O exemplo (6a), abaixo, traz uma ocorrência licenciada de elisão vocálica no PB, enquanto que o exemplo (6b) nos mostra um caso em que a segunda vogal, destacada em maiúscula, porta o acento frasal e, por isso, não permite a elisão.

- (6) a) compra uvas maduras → [com.**pru**.vas.ma.du.ras]
b) compra uvas → [com.pra.**U**.vas]
- (BISOL, 1996, p. 160)

Para Bisol (1996), o apagamento da vogal /a/ evita o choque de picos silábicos, como no exemplo (6a): em “pra” /a/ foi desagregada da sílaba e as consoantes /p/ e /r/ agregaram-se à vogal inicial da palavra seguinte, criando uma nova sílaba.

Enquanto Bisol trata dos dialetos do Sul do país, Nogueira (2007, p. 53) defende que a elisão, no dialeto paulista, também afeta a vogal [u] em posição final quando a vogal seguinte não portar o acento da frase, como podemos observar no exemplo (7b). Enquanto em (7a) o processo atende as exigências para ser aplicado, em (7b) a vogal tônica /a/ em “água” carrega o acento frasal e impossibilita sua aplicação.

- (7) a) aluno atento → [a.lu.**na**.ten.to]
b) bebo água → [be.b[u].**Á**.gua]
- (NOGUEIRA, 2007, p. 53)

O segundo processo, a ditongação, ocorre quando a vogal final de uma palavra se une à vogal inicial da palavra seguinte, formando assim um ditongo; nessa formação não há nenhum apagamento, mas há a exigência de que uma das vogais seja alta, /i/ ou /u/, e átona (BISOL, 1996, p. 160):

- (8) a) leque azul → [le.**kja**.zul]
b) menina humilde → [me.ni.**naw**.mil.de]
- (NOGUEIRA, 2007, p. 80)

Por último, há a degeminação, que ocorre, segundo Bisol, quando, entre vocábulos há o encontro de duas vogais idênticas e átonas, resultando em uma fusão (1996, p. 160):

- (9) a) menina alegre → [me.ni.**na**.le.gre]
b) leque escuro → [le.**kis**.cu.ro]

Quanto à tradução de textos musicais, para Carmo Junior (2012), o texto a ser cantado deve respeitar, principalmente, as regras musicais e, por essa razão é comum o versionista lançar mão de processos fonológicos próprios das línguas de chegada, como por exemplo, a epêntese ou apagamento de segmentos e a mudança da posição do acento tônico.

Análise do corpus

Com base no que foi apresentado nas seções anteriores, nossa proposta neste estudo é investigar se, na prática, é possível que a canção traduzida seja cantada com a mesma notação musical da canção original respeitando ou não as propriedades de acento e de processos de juntura do PB. Além disso, buscamos identificar se, no processo tradutológico, o versionista atingiu as exigências que tornam a tradução do texto musical apta ao canto. Para esse fim, e com o intuito de constatar a existência da prosódia músico-textual na tradução, analisaremos a versão do texto musical para o PB da música “Dias de Amor”, do teatro musical *Rent*, tendo como base a partitura da canção original *Seasons of Love*.

Nas sessões seguintes serão analisadas as questões rítmicas musicais e suas relações com a fala – acento e processos de juntura – e a fidelidade ou não do tradutor quanto ao conteúdo do texto musical e quanto à notação musical da canção “Dias de Amor”.

Adaptação rítmica por processos de juntura

Vimos anteriormente que, para que ocorra a correspondência nota/sílaba, o texto musical deve ser silabicamente dividido, e espera-se que cada sílaba seja atribuída a uma nota musical.

Com a partitura de *Seasons of Love* para o canto, podemos perceber que a métrica, determinada pela fórmula de compasso, é quaternária simples, ou seja, cada compasso é dividido em 4 tempos e a semínima será a nota com 1 tempo de duração:

Figura 16. *Seasons of Love*, introdução

SEASONS OF LOVE
from the Broadway Musical *Rent*

Songwriter Jonathan Larson

Fonte: elaborada pelas autoras.

Pode-se fazer a distribuição dos tempos em notas e pausas de diversas maneiras, como por exemplo:

Figura 17. Distribuição dos tempos nos compassos

Fonte: elaborada pelas autoras.

Sabemos que a *ligadura de prolongamento* une as notas, somando seus valores e fazendo com que elas correspondam a uma única sílaba. Entretanto, no trecho destacado na Figura 17 acima, temos uma *ligadura de expressão*, pois as notas ligadas são de diferentes alturas. Apesar de as ligações dessas quatro notas, compasso 33-35 (Figura 17), indicarem quatro sílabas, notamos que o letrista tende a atribuir uma só sílaba às notas com ligadura de expressão. Dessa forma, podemos dizer que, nos compassos desse exemplo cabem, respectivamente: 3 sílabas, 1 sílaba, 3 sílabas e 1 sílaba.

No texto traduzido, podemos observar a frase “de amor”, no qual a vogal final /e/ em ‘de’, realizada como [i] pelos falantes do PB, une-se à vogal inicial /a/ da palavra ‘amor’. Isso indica que as fronteiras vocálicas passaram por um processo de ditongação, correspondendo, portanto, a uma única nota:

(10) de amor → [dja.mor]

No compasso 30 há outro caso de ditongação:

Figura 18. Ditongação no compasso 30

Neste caso a última vogal de “conte” [kõtʃi] une-se ao [e] de ‘em’, formando uma única palavra de duas sílabas:

(11) conte em → [kõ.tjẽ]

No trecho que segue, apresentamos outra ocorrência de ditongação que permite que a sílaba seja associada a uma única notação musical.

Figura 19. Ditongação no compasso 46

45 tan - to que viu ou por quan - to-a - pren deu ----

Fonte: elaborada pelas autoras.

Na palavra “quanto” a vogal átona /o/, realizada como [u], une-se à vogal inicial /a/ de “aprendeu”. Temos, aí, portanto, um ditongo, que corresponde a uma única nota, resultado da junção da sílaba [to] com a vogal /a/.”

(12) quanto aprendeu → [quan.twa.pren.deu]

Entretanto, para Nogueira (2007) e Santos (2007), no dialeto paulista a vogal átona [u], em posição final de palavra, também pode ser elidida, assim como o /a/ átono na regra de elisão para os dialetos do Sul proposta por Bisol (1996), desde que as outras condições sejam também respeitadas. Assim, segundo as autoras, podemos também dizer que o processo de juntura do trecho em questão poderia ser a elisão, conforme Figura 20, abaixo.

Figura 20. Elisão

45 tan - to que viu ou por quan - t'a - pren deu ----

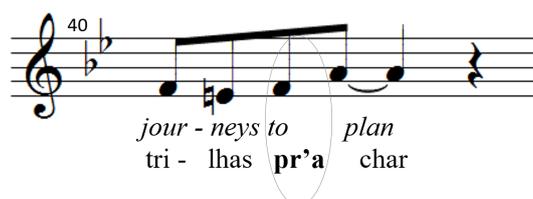
Fonte: elaborada pelas autoras.

(13) quanto aprendeu → [quan.ta.pren.deu]

Vemos, entretanto, que indiferentemente do processo aplicado, ditongação ou elisão, forma-se uma única sílaba.

No compasso 40, encontramos outra aplicação de processo de juntura que contribui para a associação da sílaba a uma única notação musical, como mostra a Figura 21:

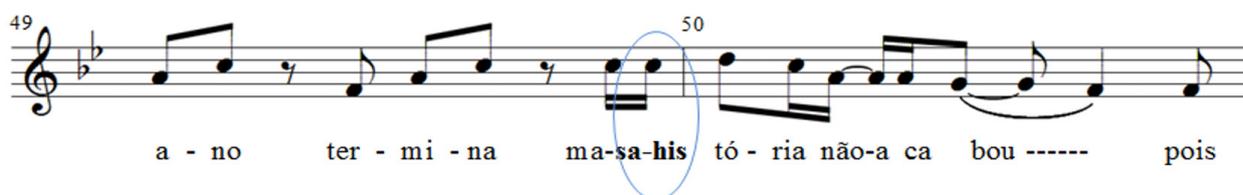
Figura 21. Degeminação



Fonte: elaborada pelas autoras.

Outro exemplo, encontrado no compasso 49, a seguir, além da silabação entre *ma[s#a]*, formando a sílaba “sa”, trata-se de um processo de ditongação para que uma correspondência entre sílaba e nota seja estabelecida:

Figura 22. Ressilabação com ditongação



Fonte: elaborada pelas autoras.

(14) mas a história → [ma.sais.tó.ria]

Percebemos, por meio das ocorrências apresentadas acima, que processos de juntura são, portanto, estratégias linguísticas a favor do versionista, pois, por meio desses processos, pode ser feita a correspondência nota/sílaba de modo que não haja alterações nos aspectos melocêntricos da canção original.

Adaptação rítmica por alteração do acento tônico

Entendemos que o acento tônico é a proeminência da sílaba na fala, uma em relação à outra e, no PB, o acento restringe-se às três últimas sílabas das palavras. Já na música, o acento métrico corresponde aos tempos fortes e fracos do compasso. No texto musical para o canto, é preferível que o acento tônico e os tempos fortes da música coincidam.

Como vimos anteriormente, a divisão métrica de *Seasons of Love* é quaternária simples (4/4), fórmula que define quais serão os tempos fortes e fracos da música e, unida ao trecho musical, pode ser representada dessa maneira:

Figura 23. Compasso 4/4, tempos forte e fraco em *Seasons of Love*

The image shows a musical staff in 4/4 time with a key signature of one flat. It covers measures 17 and 18. Measure 17 contains four notes: quarter, quarter, quarter, and quarter. Measure 18 contains four notes: quarter, quarter, quarter, and quarter. Dynamic markings are placed above the notes: *f* above the first note of measure 17, *p* above the second, *mf* above the third, and *p* above the fourth. The same sequence of dynamics (*f, p, mf, p*) is repeated in measure 18. The lyrics 'di - as em noi - tes em bei - jos em ca - fe - zi - nhos' are written below the staff, with hyphens indicating syllable boundaries. A blue dot is placed under the first note of measure 18, which is the note 'em'.

Fonte: elaborada pelas autoras.

A ligadura no compasso 18 indica uma síncopa, ou seja, ela desloca os acentos (Figura 23). Assim, essa nota que será tocada como uma só recebe o acento meio-forte.

Vejamos agora o acento tônico das palavras desses compassos:

DIas em **NOI**tes, em **BEI**jos em cafe**ZI**nhos

Sabemos que as palavras 'dias', 'noites', 'beijos' e 'cafezinhos' são paroxítonas, pois o acento tônico cai na penúltima sílaba.

Na partitura, ao fazermos a correspondência entre o acento métrico e o acento tônico, obtemos o seguinte resultado:

Figura 24. Acento métrico e acento tônico

The image shows the same musical staff as Figure 23, but with different dynamic markings and underlined letters in the lyrics. The dynamics are *f, p, mf, p* for measure 17 and *f, p, mf, p* for measure 18. The lyrics are 'DI - as em NOI - tes em BEI - jos em CA - fe - ZI - nhos'. The letters 'DI', 'NOI', 'BEI', and 'CA' are underlined in the original image, indicating the tonic accent. A blue dot is placed under the first note of measure 18, which is the note 'em'.

Fonte: elaborada pelas autoras.

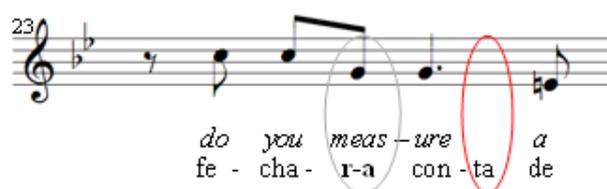
Podemos perceber que o acento tônico e os tempos fortes da música são correspondentes em **DI**, **NOI** e **BEI**, como esperado. Na palavra “cafezinho”, porém, há um deslocamento do acento tônico de “ca**FÉ**” para “**CA**fé” a fim de que corresponda ao tempo forte da música, sendo perceptível, ainda que menos intenso. Em versões musicais, a alteração do acento é comum, pois o texto musical traduzido está submetido à partitura do texto musical original, porém essa alteração deve ser o último recurso ao qual o versionista deve recorrer, pois, segundo Cagliari (2012, p. 29), quando a música cantada viola a estrutura da língua, pode causar estranheza para o ouvinte.

Alteração da notação musical

Vimos anteriormente que, na tradução do texto musical para o canto, o versionista deve dar importância às questões músico-textuais das canções e deve respeitar a notação musical da canção original, pois esta não deve ser alterada. Tais questões já foram previamente analisadas neste estudo.

Em “Dias de Amor”, entretanto, encontramos casos de alterações na partitura, como mostra o trecho a seguir.

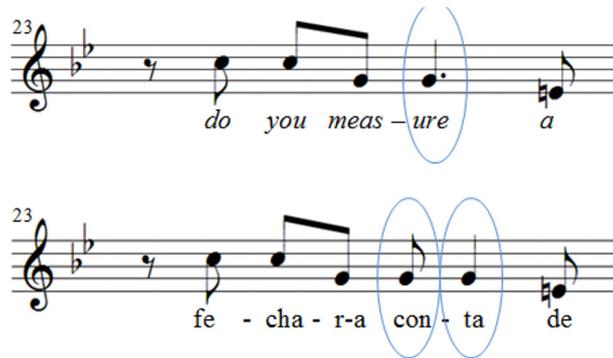
Figura 25. Ausência da nota



Fonte: elaborada pelas autoras.

Nesse compasso, é visível a ausência de uma nota que corresponda à sílaba “ta”. Para resolver essa falta de correspondência, a versionista precisou alterar a partitura original. O ponto de aumento do tempo da semínima, a nota que faz correspondência com a sílaba “con”, foi transformado em colcheia, pois, teoricamente, ambos, nesse caso ponto e colcheia, possuem o mesmo valor de tempo de duração ($\text{♩} = \text{♪} + \text{♪}$), apresentando o seguinte resultado:

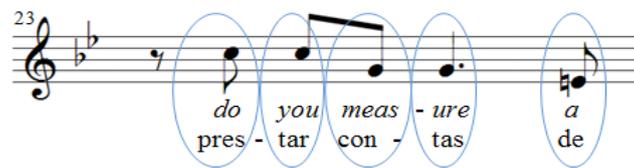
Figura 26. Alteração da notação musical 1



Fonte: elaborada pelas autoras.

Uma solução à vista disso seria o uso da expressão “prestar contas”, pois seria fiel à logocentria e, desse modo, excluiria a necessidade do uso do artigo “a”, reduzindo o número de sílabas de forma a haver coincidência com as notas musicais da partitura original:

Figura 27. Solução para a alteração da notação musical 1



Fonte: elaborada pelas autoras.

Contudo, há casos em que a informação dada na música original não pode ser adaptada, muitas vezes por se tratar de uma expressão consagrada cuja alteração não teria boa aceitação do público; nestes casos, ao versionista não resta outra opção senão traduzir o texto musical literalmente. Em *Seasons of Love*, o verso inicial refere-se ao tempo em minutos contido em um ano, um fato que não pode ser alterado na tradução. Entretanto, ao analisarmos a correspondência silábica entre as expressões, temos:

five hun.dred twen.ty-five thous.and six hun.dred min.utes – 8 palavras 13 sílabas
qui.nhen.to.se vin.te cin.co mil seis.cen.tos mi.nu.tos – 7 palavras 15 sílabas

Silabicamente, a informação original e a traduzida não coincidem. Devido a isso, a tradução para o PB não encaixa na notação musical da canção original. Nesse caso, vemos

que a versionista lança mão de uma modificação na notação musical para que, no canto, as palavras não causem estranheza ao ouvinte:

Figura 28. Alteração da notação musical 2

The figure displays two musical staves in G major (one sharp) and 4/4 time. The first staff shows the original English lyrics: "five hun-dred twen-ty five thous-and six hun-dred min - utes". The second staff shows the Portuguese translation: "qui - nhe-to-s-e vin-te cin-co mil seis - cen - tos mi - nu - tos". The notation is identical in both staves, but the syllable division in the lyrics is altered to fit the musical rhythm. For example, "twen-ty" is split as "twen-ty" and "fif-ty" as "fif-ty". The measure numbers 9 and 10 are indicated above the staves.

Fonte: elaborada pelas autoras.

O exemplo acima foi, portanto, uma exceção às regras da tradução de músicas para o canto, pois no conflito entre o logocêntrico e o melocêntrico, o literalismo foi de maior relevância para que informação fosse precisamente transmitida.

Considerações finais

O objetivo geral desta pesquisa foi analisar o processo tradutológico presente nas versões de músicas para o canto no teatro musical, averiguando de quais estratégias o versionista lança mão para que a tradução seja tanto logocêntrica como melocêntrica. Conforme Frazon (2008), a música traduzida cantável é aquela que atinge todos os níveis prosódicos músico-textuais, sendo eles a correspondência silábica, a correspondência entre o acento tônico e os tempos fortes da música, a combinação poética e a combinação semântica.

Com base nesses princípios, percebemos que a canção analisada cumpriu com as exigências que a tornam cantável: da mesma forma que a canção em PB pode ser acompanhada junto à melodia da canção em inglês, a música traduzida não interfere nos efeitos de sentido da música original. Porém, na partitura, que não deve ser alterada, foram encontradas duas mudanças quando a tradução precisou ser literal e não houve a possibilidade de uma adaptação.

Notamos também que aspectos da fonologia do PB podem ser uma ferramenta no trabalho do versionista. Tanto a alteração do acento tônico quanto processos de juntu-

ra foram recursos utilizados para que houvesse a correspondência métrica entre texto e música, com o intuito de que o texto musical traduzido possa ser cantado juntamente à partitura original.

Desse modo, concluímos que o versionista tende a manter o melocentrismo da canção original e que suas traduções são adaptações que, embora com outras palavras, criem o mesmo efeito de sentido a fim de que a música original e a música traduzida soem como uma só.

Referências

- BENNETT, R. *Elementos Básicos da Música*. Rio de Janeiro: Jorge Zahar Editor, 1987.
- BENWARD, B.; SAKER, M. *Music in Theory and Practice*. 8. ed. New York: McGraw-Hill Education, 2009. v. 1.
- BISOL, L. O Sândi e a Ressilabação. *Letras de Hoje*, Porto alegre, v. 31, n. 2, p. 159-168, jun. 1996.
- CAGLIARI, L. C. Línguas de Ritmo Silábico. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 20, n. 2, p. 23-58, jul./dez. 2012.
- CARMO JÚNIOR, J. R. do. Sobre a Gramática da Palavra Cantada. *Cadernos de Estudos Linguísticos*, Campinas, v. 54, n. 2, p. 205-222, jul./dez. 2012.
- CARVALHO JÚNIOR, A. D. de. *Análise de padrões Musicais Rítmicos e Melódicos Utilizando o Algoritmo de Predição por Correspondência Parcial*. 92p. Dissertação (Mestrado em Sistemas de Computação) – Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, 2011.
- COLLISCHONN, G. *O acento em Português*. In: BISOL, L. Introdução a estudos de fonologia do português brasileiro. 3. ed. Porto Alegre: EDIPUCRS, 2001, p. 125-131.
- FRAZON, J. Choices in Song Translation. *The Translator*, London, v. 14, n. 2, p. 373-399, out. 2008.
- HASTY, C. F. *Meter as Rhythm*. New York: Oxford University Press, 1997.
- LACERDA, O. *Compêndio de Teoria Elementar da Música*. 3. ed. São Paulo: Ricordi Brasileira, 1967.
- LARSON, J. *Rent*. New York: Music Theatre International, 1994. 1 partitura, 8 p. Piano.
- MED, B. *Ritmo*. 3. ed. Brasília: MusiMed, 1984.
- MEDEIROS, B. R. de. Ritmo na língua e na música: o elo possível. *Música em Perspectiva*, Curitiba, v. 2, n. 2, p. 45-63, out. 2009.
- MIGLIORINI, L. M. de Q.; MASSINI-CAGLIARI, G. Sobre o ritmo do português brasileiro: evidências de um padrão acentual. *Revista Virtual de Estudos da Linguagem*, São Paulo, v. 8, n. 15, p. 310-328, 2010. Disponível em: <http://hdl.handle.net/11449/124843>. Acesso em: 22 fev. 2017.

NOGUEIRA, M. V. *Aspectos Segmentais dos Processos de Sândi Vocálico Externo no Falar de São Paulo*. 144 p. Dissertação (Mestrado em Linguística) – Departamento de Semiótica e Linguística Geral, Universidade de São Paulo, São Paulo, 2007.

SCHWARTZ, S. *A New Musical Wicked*. Minnesota: Hal Leonard, 2003. 2 partituras, 12 p. Piano/ Vocal.

TENANI, L. E. *Domínios Prosódicos no Português*. 377p. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, São Paulo, 2002.



Data de submissão: 15/03/2018

Data de aceite: 19/11/2018

HOW PROSODY CONSTITUTES THE ACTIONS OF FORMULATIONS IN COURTROOMS¹

COMO A PROSÓDIA CONSTITUI AÇÕES DE FORMULAÇÕES NA CORTE

Ana Cristina Ostermann | [Lattes](#) | ACO@unisin.br
Universidade do Vale do Rio dos Sinos

Daniela Negraes Pinheiro Andrade | [Lattes](#) | danielanegraes23@gmail.com
Universidade do Vale do Rio dos Sinos

Minéia Frezza | [Lattes](#) | mineiafrezza@hotmail.com
Universidade do Vale do Rio dos Sinos

Resumo: Com base nas perspectivas teóricas e metodológicas da Linguística Interacional e da Análise da Conversa, este estudo investiga como o traço prosódico tom contribui para a formação e a atribuição de sentido às ações em audiências de instrução, em particular, na prática de formulação. A análise revela que variações de tom na prática de perguntar estão intrinsecamente relacionadas às diferentes ações realizadas local e situadamente: checar entendimento e desafiar o interlocutor.

Palavras-chave: Análise da conversa; Linguística interacional; Prosódia; Formulações.

Abstract: Based on the theoretical and methodological perspectives of Conversation Analysis and Interactional Linguistics, this study investigates how the prosodic feature pitch contributes to the formation and ascription of actions in court hearings, in particular, in practices of formulation. The analysis reveals that variations of pitch in the practice of formulating are intrinsically related to the different locally and situated actions performed: checking understandings and challenging the interlocutor.

Key words: Conversation Analysis; Interactional Linguistics; Prosody; Formulations.

¹ A modified and extended version of the discussion presented here was published in Brazilian Portuguese in <http://ken.pucsp.br/delta/article/viewFile/20030/22218>.

1 Introduction

Based on the perspectives of Conversation Analysis and Interactional Linguistics, this paper analyzes the prosodic features of courtroom interactions that present the interactional practice of formulating conversation. Formulations are described as an interactional practice that allows participants to perform different actions insofar as they enable “[...]co-participants to settle on one of many possibilities of what they have been saying” (HERITAGE; WATSON, 1979, p. 123). In this sense, a formulation “constitutes a shift from the exchange of information and particulars about the issues under negotiation to a review of ‘where we/you are’” (WALKER, 1995, p. 104) which orients the interlocutor to the implementation of certain subsequent actions (HERITAGE; WATSON, 1979). Such orientation may be observed and described insofar as the production of a formulation opens the first-pair part of an adjacent pair (e.g. checking understanding) and makes the interlocutor responsible for providing the second-pair part of the adjacency pair (e.g. confirming/disconfirming that understanding) to close that conversation sequence.

Occurrences of formulation are not abundantly observed in ordinary talk but are more likely to happen in institutional talk – e.g. court hearings, police interrogations, decision-making meetings – due to “institutionalized distribution of rights to formulate” (HERITAGE; WATSON, 1979, 150). In this paper, we will specifically discuss two types of actions performed through the practice of formulating: (a) checking understanding and (b) challenging the interlocutor. The second action (b) may be observed when the formulation done as the first-pair part requires that the interlocutor provides an account as the second-pair part of the adjacency pair (SACKS, 1992). While type (a), checking understanding, is one of the most recurrent actions implemented with the use of formulations, even in ordinary talk, challenging the interlocutor seems to be much less common. However, as providing accounts is one of the main characteristics of interactions in hearing audiences, it might not be surprising to find a collection of them in such context which, *per se*, justifies this study. In addition, the investigation of the role prosodic features play in the practice of formulating should contribute for enlarging the studies of naturally occurring interactions in Brazilian Portuguese, in particular from the interactional linguistic and conversation analytic perspectives.

2 Prosody in naturally occurring interactions

When Sacks, Schegloff and Jefferson (1974) proposed Conversation Analysis (CA) methodology to describe social order based on the empirical investigation of

naturally occurring interactions, they called attention to the importance of phenomena such as hesitation, pauses, laughter, sound lengthening, inspiration, expiration, speech cut-offs, pitch, and intonation, to the local and situated co-construction of interactional actions. Since the very beginning of CA studies, then, prosody has been understood as one of the resources participants rely on to manage, negotiate, and attribute meaning to social actions in interaction (COUPER-KUHLEN; SELTING, 1996; LEVINSON, 2013).

In order to assist CA researchers to understand in which ways prosody operates in action formation and ascription in naturally occurring interactions, Emanuel Schegloff (1998) suggests two analytical approaches: (a) focusing the research on comparative analysis of prosodic phenomena on occasions of spontaneous talk-in-interaction events and other contexts (e.g. reading); (b) focusing the research on conversational events in order to investigate prosody as one of the elements which contribute to its organization. Our research follows the second analytical approach (b). Hence, we take prosody as a set of resources which contribute to the production of talk-in-interaction, focusing on how it contributes to meaning ascription of certain actions (i.e. checking understanding and challenging the interlocutor) in formulations produced in the courtroom interactions.

3 Methodology

This study follows a qualitative analytical approach, with the data analysis discussion being supported by a combination of the theoretical and methodological perspectives of Conversation Analysis (CA) and Interactional Linguistics (IL). As CA was first developed in a sociological intellectual environment, several studies in the area have recurrently neglected some linguistic aspects of talk such as prosody. According to Selting (2010, p. 6), “there is no spoken language without prosody, and disregarding prosody in the study of spoken language means disregarding an integral part of spoken language itself.” By aiming at fulfilling this gap – on what concerns interactional linguistic studies of Brazilian Portuguese – we rely on the theoretical and methodological approach of Interactional Linguistics which investigates how prosody – as a suprasegmental aspect – may contribute to the ascription of meaning to specific interactional actions in naturally occurring conversation.

Following Conversation Analysis premises to generate data, the methodological procedures undertaken were: (a) observing and audio recording 59 hearing audiences; (b) transcribing all recordings; (c) continuously listening to the recordings together

with finely revising the transcripts; (d) building a collection of recurring patterns of formulations done by the institutional representative; (e) sequentially analyzing the actions implemented by means of formulations; (f) by using the software Praat, analyzing the prosodic features of the turns of talk in which formulations were produced; and (g) investigating the role of prosody in the accomplishment of the specific actions in formulations.

Part of CA and IL research comprises transcribing data based on a set of conventions that intend to allow readers to understand what was said beyond the reproduction of standard orthography. The convention system chosen to transcribe data was the Jeffersonian one, which represents “a magical engagement with the lived moment-by-moment features of interaction” (HEPBURN; BOLDEN, 2017), as represented in Figure 1.

Figure 1. Jeffersonian conventions

Transcription conventions based on Jefferson (1984)

(1.8)	Sound
(.)	Micro pause
=	Latched talk
[text]	Overlap
,	Continuing intonation
↑text	Rising intonation immediately following the arrow
↓text	Falling intonation immediately following the arrow
.	The whole turn is produced in falling intonation
?	The whole turn is produced in rising intonation
-	Abrupt cutoff
:::	Sound lengthening
>text<	Quicker pace than the surrounding talk
<text>	Slower pace than the surrounding talk
TEXT	Louder than the surrounding talk
°text°	Quieter than the surrounding talk
text	Emphasis
(text)	Transcriptionist doubt
xxxx	Inaudible talk
((Texto.))	Vocalizations and other details of the conversational scene
hhh hahahahehihi	Laughter
.hhh	Audible inhale
hhh	Audible exhale

Source: Jefferson (1984, p. ix-xvi).

As for analyzing prosody, to build collections, we focused on the instances in which the prosodic patterns revealed in the formulations produced different actions. The selected extracts focus on one of the institutional representative’s talk: a female judge.

Such decision was grounded on Szczepiek Reed's (2010) orientations for conducting prosodic analyses: firstly, by comparing extracts of talk produced by the same person and, secondly, by comparing extracts of talk produced by different people engaged in a conversation. From our corpus of 59 interactions, we selected two instances in which the practice of formulations was constituted as a means of checking understanding, and two formulations used to challenge the interlocutor. For this the purposes of this paper, we discuss how these two types of formulation differ in terms of prosody by showing one occurrence from each collection.

The excerpts cited in this article are transcribed in three lines: for each numbered line, there is the transcription in Brazilian Portuguese, the original language of the data. The second line presents a word-by-word English gloss, and the third one, a free English translation.

4 Data analysis

Excerpt 1 shows how the action of checking understanding via formulation also accomplishes the action of putting information "on record" – one of the legal settings' institutional goals (ANDRADE, 2010). The excerpt under scrutiny is part of a hearing audience in which the judge (JUD) interrogates the defendant (DEF) about his involvement with the crime of drug dealing (crack). In the previous turns of talk (not shown here), the defendant claims he is not involved with the alleged crime and points out his son as the plaintiff. The defendant also claims his son was a drug addict at the time the crime was perpetrated.

Excerpt 1 [FCSLAIRJ0413SET08]

(J= Judge, D= Defendant)

- 12 J: mas que história é essa da sua filha
but what story is this-FEM of the-FEM your-FEM daughter
but what about this story of your daughter
- 13 a sua filha também tá pre:sa?
the your-FEM daughter also is arrested-FEM
your daughter is she also in jail?
- 14 D: tá:: °tá envolvida com° tráfico também
is is involved-FEM with traffic too
yeah she's also involved with drug dealing
- 15 J: **tá então a sua fi::lha então é essa daqui:**
okay so the-FEM your-FEM daughter so is this-FEM of he:re
okay so your daughter is this one here
- 16 D: é:: é a irmã do joãzi::nho °a minha filha°
is is the-FEM sister of the-MAS joão-DIM the-FEM my-FEM daughter
yeah she's johnny's sister my-FEM daughter
- 17 (4.7)
- 18 J: tá mas e o::: (5.4) quem é essa fi::lha
okay but and the-MAS who is this-FEM daughter
okay but and who's this daughter of yours

Based on the defendant's testimony and on the dossier the judge holds in her hands,² the judge orients to the defendant's daughter's involvement in the legal process being analyzed: "but what about this story of your daughter is she also in jail?" (line 12-13). The defendant responds confirming that information.

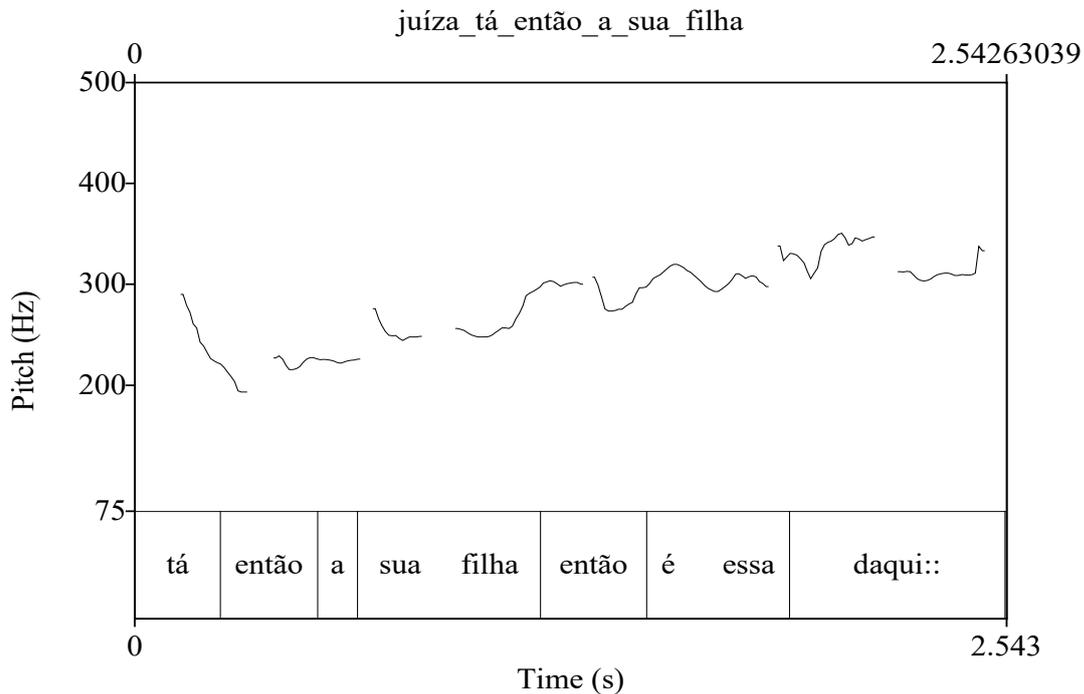
In addition, the defendant offers more information than what he had been requested. He provides the judge the reason why his daughter is in jail, that is, she is involved with drug dealing. The judge, then, formulates her understanding concerning the defendant's daughter's identity: "okay so your daughter is this one here" (line 15).³ The defendant does not treat the judge's turn as an assertion, but as request for confirmation of an understanding instead, as he provides the second-pair part of the adjacency-pair initiated by judge when she formulates her understanding concerning the daughter's identity. In so doing, he confirms his daughter's identity (line 16). The action of checking understanding performed via formulation may be observed (a) in the sequence in which the formulation is in the conversation (line 15), given the interlocutor's adjacent confirmation (line 16).

To analyze the pitch range when such formulation is produced, it is important to inform that the judge's pitch range oscillates between 140Hz and 360Hz and her average range is approximately 270Hz. The fundamental frequency contour of the turn of talk where she produces a formulation shows that when she produces "okay" (*tá*), in the beginning of the turn, the pitch range reaches 288Hz. Later, when she produces "so" (*então*) – a discourse marker commonly used to express logical deduction such as "if p, so q", and commonly used in formulations –, the pitch range falls to 228Hz. Then, the pitch range increasingly rises and reaches its peak when the judge produces the deictic "here" (*daqui*) (347Hz). In addition, we can notice that word lengthening presents rising pitch range as well.

² Information based on field notes.

³ Field notes show that in line 15, when the judge produces "this one", she is pointing to the dossier she is handling.

Figure 2. Fundamental frequency contour of “so your daughter is this one here”



The fundamental frequency contour shows that the pitch of the formulation: (a) presents a continuously rising curve after the “so” (*então*); and (b) reaches its peak when the judge expresses her understanding about who the deponent’s daughter is – that is, “this one here” (*essa daqui:*) and nobody else who had been mentioned up to that moment. This analysis shows that the prosodic features produced by the judge during a formulation that checks understanding are different from the ones that challenge the interlocutor, as discussed later in this paper. The challenging character of certain turns is not given aprioristically but evidenced by the orientation of the next speaker; i.e. based on the next speaker’s contribution to the ongoing interaction. One of the most concrete emic evidence which can work as proof procedure that the next speaker has understood certain turn as a “challenge” is their provision of accounts in their next turn. We show next how this happens.

In Excerpt 2, the legal case refers to a woman who claims to restore the custody of her child that, after her parents’ divorce, got under the father’s responsibility, who is the deponent’s former husband. The deponent (DEP), who is the child’s mother, alleges to have been forced to get divorced because of domestic violence due to groundless suspicion that she had betrayed her husband. The following interaction happens after the judge (JUD) inquires the deponent about her current partner.

Excerpt 2 [FCSLAIRJ0413SET08]

(J= Judge, D= Deponent)

- 8 J: quanto tempo a senhora morou com esse
how long you lived with this-MAS
how long have you lived with this
- 9 novo [°companheiro°.]
new-MAS partner-MAS
new partner
- 10 D: [um ano e nove] meses.
a-MAS year and nine months
a year and nine months
- 11 (1.9)
- 12 J: com ↑qual
with what
with whom
- 13 (.)
- 14 com esse ↑úl[:ti]mo?
with this-MAS last-MAS
with this last one
- 15 D: [é:]
is
yes
- 16 (14.6) ((The judge puts the deponent's testimony on record by dictating it to her secretary.))
- 17 J: >bom então:< (0.2) pela minhas contas aqui
good so by the-FEM my-FEM calculations here
well so if I'm correct here
- 18 se foi em ↑mar:ço
if was in march
if it was in march
- 19 (1.2)
- 20 J: **então a senhora ↑lo:go depois já se juntou**
so the-FEM lady soon after already yourself joined
so soon after you already moved in
- 21 **com [↑ou:tro]?**
with other
with other
- 22 D: [↑não:] (°porque°) (0.5) >se conhe↑cemo< há nove meses
no because ourselves met has been nine months
no because we met nine months ago
- 23 D: (.) **nós fomos morá junto mes:mo (faz) uns seis mês e pouco.**
we went to live together indeed makes some six months and a little
we moved in together it's been about six months indeed

The judge requests the time length the deponent has lived with her new partner, which is responded to in line 10. The judge then produces a formulation (line 17-20), which is actually “held up” during its production (see gap in line 19). The design of such turn of talk (lines 17-20) is particularly interesting because it shows the reflexive procedure (GARFINKEL, 1967) that grounds the judge’s conclusion socially shared by means of a formulation in line 20. In fact, the judge shows that her “math reasoning”

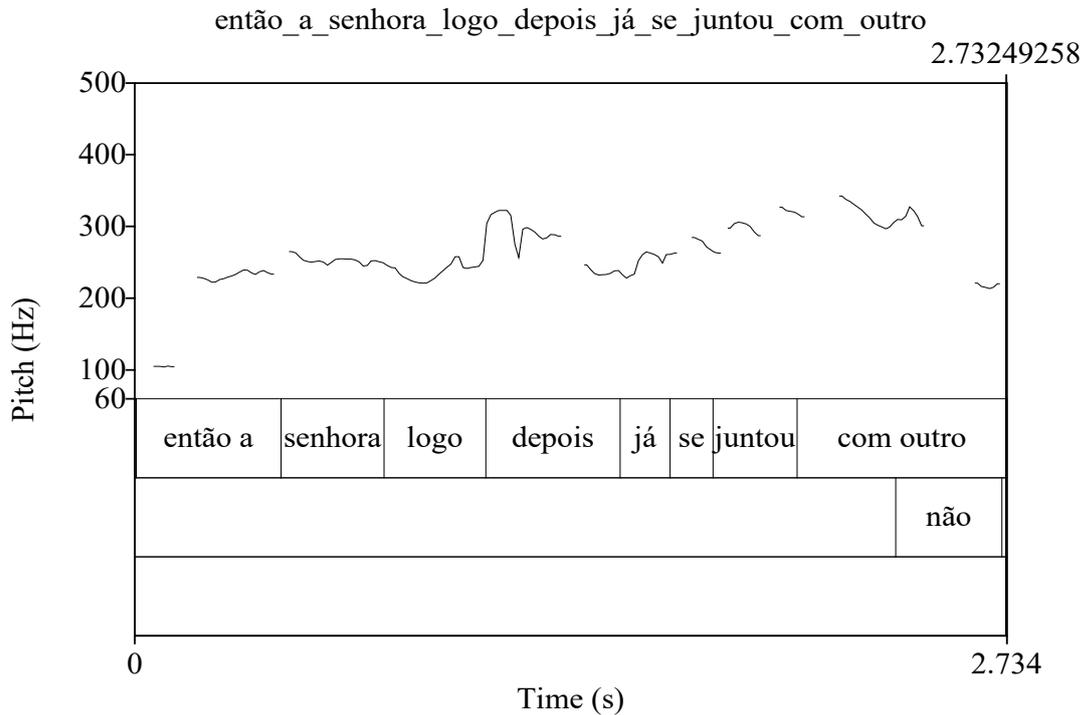
had already started in line 17. This explains the absence of talk in line 19 – somewhat longer than others in this interaction – but which is possibly a result of time the judge needs to calculate the amount of time it took the deponent to move in with another man after the break up so – presented by means of a formulation. Sacks (1992) illuminates the analysis of the second part of the formulation of the judge.

In addition to the demand that conversation analysts build their analysis based on the sequential organization formed by the turns of talk, Schegloff (2007) claims that it is necessary to consider the various layers that operate on talk organization. He claims that the organizational layers work on the sequentiality of the adjacency pairs, on the macrostructure of conversation and on the topics made relevant in the conversation. In terms of this last aspect, Sacks (1992) says that whenever it is not possible to end up a topic in one adjacency pair, the topic tends to be expanded so as to connect what we had been saying previously to what we are saying now, even if we are not talking about the same thing now. The formulation in Excerpt 2 seems to request an elaboration from the deponent about her relationship with her current interactional partner. Thus, the formulation of the judge requests an account and challenges the deponent about the truth of facts that her former husband's suspiciousness was groundless.

As our analytic object consists of an adjacency pair, it is crucial to check its second pair-part – the deponent's answer (lines 22-23) – that is produced in overlap with the end of the judge's formulation. The deponent shows her orientation to provide an account as she explains that she had first met her current partner and only after a certain amount of time had moved in with him. The account provided by the deponent can be interpreted as evidence that the deponent had oriented herself to the judge's talk as a potential "moral" challenge of her being a "decent woman" and of her version of the facts.

In terms of prosody, Figure 3 represents that the production of this formulation differs from the one analyzed in Excerpt 1.

Figure 3. Fundamental frequency contour of “so you soon after moved in with another”.



The connector “so” (*então*) is produced with rising pitch range (starting at 105Hz and ending at 234Hz). Then, the pitch remains high along part of the production of the pronoun “you” (*senhora*). The adverb “soon” (*logo*) reaches a much higher pitch range (starting at 242Hz and ending at 322Hz), especially when compared to the last syllable of the previous word. Soon after the pitch range falls again and reaches a new peak at the moment the judge produces the word “another” (*outro*) (starting at 342Hz and ending at 223Hz). Therefore, the words “soon” and “another” are both produced with higher pitch, which, seen from a semantic perspective, locally and sequentially speaking, constitute the focus of the challenge done via this formulation. In addition to these lexical choices, the prosody the judge uses to produce these words contributes to the construction of the turn as potentially implementing the action of challenging. To justify our analysis, we rely on Couper-Kuhlen and Selting’s claim (1996, p. 21) concerning the “pragmatic meaning” a certain intonation can bear in situated contexts, by claiming that the prosody used in the turn where the formulation is produced by the judge (in Excerpt 2) makes a negative type-conforming response (RAYMOND, 2003) preferred. Our analytical argument, thus, underpins the statement of Couper-Kuhlen and Selting (1996, p. 21), in which they affirm that “prosodic contextualization cues are not referential, but indexical, signs” and, therefore, are imperative to the analytic studies which are embedded with the

description of the social machinery of language. The formulation contained in Excerpt 2 has a prosodic production that can operate to: (a) transform the polarity of the preferred response (i.e., confirmation or disconfirmation), and (b) make an account relevant in the next turn. To sum up, the pragmatic meaning imprinted by the judge's prosody onto her turn of talk where the formulation is produced works to challenge the deponent. The deponent's account (line 25-26) becomes the next turn proof procedure that the deponent does indeed ascribe the action of a request for an account to the judge's formulation.

5 Conclusions

This study reveals how the actions of information checking and of challenging are implemented also using different prosodic features. Based on Levinson's (2013) proposal, we analyzed prosody as part of a linguistic and formal layer combined to a sequential and contextual analysis to describe how the process of designing and ascribing meaning to the actions implemented by the practice of formulating takes place.

Importantly, the lexical choices that represent the focus of judge's formulations (either as checking understanding or challenging the information provided) are produced prominently. As a result, the next speaker is led to interpret the formulations based on the pragmatic meaning indicated by their marked prosody.

Therefore, we claim that the prosodic characteristics observed in the judge's formulations work as another layer in the design and meaning ascription of the actions being implemented (LEVINSON, 2013): formulating to check understanding and formulating to request for accounts (the latter working some type of confrontation). Two factors support such claim: (a) the clue that the actions are based on the emphasis of the terms to be confirmed or disconfirmed by the next speakers, and, (b) by the next speakers' orientations to provide the second pair-part displaying how they ascribe meaning to the expectations generated by prosody used in the actions implemented in the first pair-part.

References

ANDRADE, D. *O uso de referentes pessoal e de lugar e o uso de formulações em interrogatórios na corte*, São Leopoldo. Master's Dissertation. Universidade do Vale do Rio dos Sinos – UNISINOS, 2010.

COUPER-KUHLEN, E.; SELTING M. (Ed.). *Prosody in conversation: interactional studies*. New York: Cambridge University Press, 1996.

GARFINKEL, H. *Studies in Ethnomethodology*. Englewood Cliffs, 1967.

HEPBURN, A.; BOLDEN, G. *Transcribing for social research*. Los Angeles: Sage Publications, 2017.

HERITAGE, J.; WATSON, R. Formulations as conversational objects. In: PSATHAS, G. (Ed.), *Everyday language: studies in ethnomethodology*. Mahwah: Lawrence Erlbaum Associates, 1979, p. 123-162.

JEFFERSON, G. Transcript notation. In: ATKINSON, J.; HERITAGE, J. *Structures of social action: studies in conversation analysis*. New York: Cambridge University Press, 1984. p. ix-xvi.

LEVINSON, S. Action formation and ascription. In: SIDNELL, J.; STIVERS, T. (Ed.). *The Handbook of Conversation Analysis*. Massachusetts: Wiley-Blackwell, 2013, p. 103-130.

RAYMOND, G. Grammar and social organization: yes/no interrogatives and the structure of responding. *American Sociological Review*, v. 68, p. 939-967, 2003.

SACKS, H. *Lectures on conversation*. Oxford: Blackwell, v. 1 e v. 2, 1992.

_____; SCHEGLOFF, E.; JEFFERSON, G. The simplest systematics for turn-taking in conversation". *Language*, v. 50, n. 4, p. 696-735, 1974 .

SCHEGLOFF, E. Reflections on studying prosody in talk-in-interaction, *Language and Speech*, v. 42, n. 3-4, p. 235-263, 1998.

_____. *Sequence organization in interaction: a primer in Conversation Analysis*. Cambridge: Cambridge University Press, v. 1, 2007.

SELTING, M. Prosody in interaction: State of the art. In: BARTH-WEINGARTEN, D.; REBER, E.; SELTING, M. (Ed.). *Prosody in interaction*. Amsterdam: Benjamins, 2010, p. 3-40.

SZCZEPEK REED, B. *Analysing conversation: an introduction to prosody*. New York: Palgrave Macmillan, 2010.

WALKER, E. Making a bid for change: formulations in union/management negotiations. In: FIRTH, A. (Ed.). *The discourse of negotiation: studies of language in the workplace*. Oxford: Pergamon, 1995, p. 101-140.



Data de submissão: 18/01/2018

Data de aceite: 18/09/2018

FRASEAMENTO PROSÓDICO DE ESTRUTURAS PARENTÉTICAS EM DADOS DE LEITURA NO PORTUGUÊS DO RIO DE JANEIRO

PROSODIC PHRASING OF PARENTHETICAL STRUCTURES
IN READING DATA OF BRAZILIAN PORTUGUESE FROM RIO DE JANEIRO

Vivian Borges Paixão | [Lattes](#) | vpaixao91@gmail.com

Universidade Federal do Rio de Janeiro

Carolina Ribeiro Serra | [Lattes](#) | carolserrafuj@gmail.com

Universidade Federal do Rio de Janeiro

Resumo: O presente artigo focaliza a realização de estruturas parentéticas no português do Brasil (PB), a partir de um ponto de vista fonológico. Partindo da literatura sobre o tema – não só no escopo da fonologia, mas também abarcando a interface com a sintaxe e os estudos do texto –, realizou-se a análise acústica e entoacional de 568 sentenças lidas por informantes cariocas do sexo feminino (42 sentenças x 15 falantes, excluídos 62 dados), a partir do aparato teórico-metodológico da Teoria Autossegmental e Métrica da Entoação e da Fonologia Prosódica, a fim de contribuir para a descrição da gramática entoacional do PB, no que se refere ao fraseamento prosódico. A análise considerou a ocorrência e duração de pausas silenciosas antes e depois da parentética, além dos movimentos tonais no interior dos três sintagmas entoacionais (IPs) ideais que constituem cada enunciado. Confirmaram-se duas das hipóteses iniciais: (i) a de que a estrutura parentética é realizada, no PB, como um IP independente e (ii) a de que o fraseamento prosódico da parentética é sensível ao tamanho do primeiro IP do enunciado. Os resultados revelaram diferenças interessantes em relação àqueles obtidos por outros pesquisadores para o português europeu (PE): no PB, os dados indicam serem mais frequentes as pausas no final da parentética e os contornos descendentes no primeiro e no segundo IPs, enquanto no PE predominam as pausas no início da parentética e os contornos ascendentes.

Palavras-chave: Parentéticas; Fraseamento prosódico; Entoação; Pausa; Português do Brasil.

Abstract: This article focuses on the realization of parenthetical structures in Brazilian Portuguese (BP), from a phonological point of view. We departed from the literature review, not only in the scope of phonology, but also in the interface with syntax and text studies, and we did acoustic and intonational analysis of 568 sentences read out (42 sentences x 15 speakers, excluding 62 data), from the theoretical and methodological framework of the Autosegmental-Metrical Theory of Entonation and of the Prosodic Phonology, in order to contribute to the description of the intonational grammar of BP, with regard to prosodic phrasing. The analysis considered the occurrence and duration of silent pauses before and after the parentheticals, as well as the tonal movements within the three ideal intonational phrases (IPs) that constitute each utterance. Two of the initial hypotheses were confirmed: (i) that the parenthetical structure is performed in BP as an independent IP and (ii) that the prosodic phrasing of the parenthetical is sensitive to the size of the first IP of the utterance. The results revealed interesting differences in relation to those obtained by other researchers for European Portuguese (EP): in BP, our data indicate that pauses on the left of the parenthetical are more frequent than pauses on its right. Also, we found out that the descending contours in the first and second IPs are predominant. In EP, differently, literature indicates that pauses at the beginning of parentheticals predominate, as well as ascending contours in the first and second IPs.

Keywords: Parentheticals; Prosodic phrasing; Intonation; Pause; Brazilian Portuguese.

1 Apresentação do objeto de estudo

Este artigo tem como objetivo investigar de que forma se dá o fraseamento prosódico de estruturas parentéticas no português do Brasil (PB) – fala carioca, em dados de leitura de frases, a partir de uma abordagem fonológica. A hipótese testada é a de que há um fraseamento *default* dessas estruturas, no PB, no sentido de constituírem um sintagma entoacional (IP) independente dos elementos à esquerda e à direita da parentética; e de que o tamanho do IP à esquerda é um fator importante para essa delimitação: quanto maior o elemento à esquerda, mais robustas as fronteiras prosódicas entre esse elemento e a parentética.

O primeiro passo no sentido de caracterizar prosodicamente as parentéticas consiste em conceituar essas estruturas, ou seja, apresentar o termo usado para definir nosso objeto de estudo. Para tanto, é fundamental estabelecer diálogo com trabalhos na área da linguística textual que se debruçam sobre o mesmo tema. A própria denominação “pa-

rentéticas” remete a uma perspectiva textual-discursiva, e a maior parte da literatura que busca caracterizar tais estruturas, em especial no PB, o faz desse ponto de vista.

Os trabalhos de Nicole Dehé são, possivelmente, aqueles de maior relevância no cenário internacional no que tange ao tópico das inserções parentéticas. Em seu livro *Parentheticals in spoken English: the syntax-prosody relation* (DEHÉ, 2014), a autora faz uma exaustiva revisão da bibliografia sobre o tema, apresentando um levantamento das várias definições oferecidas pela literatura. Busca, ainda, a interseção entre as ideias apresentadas por cada um dos autores estudados, afirmando, sinteticamente, que todas as definições apontadas por eles

têm em comum o fato de que uma parentética é considerada uma entidade linguística que é linearmente integrada em outra estrutura linguística, mas não é relacionada ao material linguístico circundante de uma forma ou de outra, i. e., em termos de estrutura sintática, significado semântico e/ou entoação. (DEHÉ, 2014, p. 1, tradução nossa).¹

Essa definição constitui um denominador comum entre aquelas propostas por outros autores, como Burton-Roberts (2006, p. 179), que afirma que uma parentética (P) é “uma expressão sobre a qual se pode afirmar que, sendo, em certo sentido, ‘hospedada’ por outra expressão (H), não traz qualquer contribuição para a estrutura de H”², ou seja, P e H são estruturalmente não relacionadas.

No âmbito dos estudos do discurso, inclusive do PB, há alguns trabalhos que tratam das construções parentéticas, com destaque para os de Jubran (1996a, 1996b, 1999, 2002, 2006) e Jubran et al. (1993, 1996), que integram os estudos do grupo de Organização Textual-Interativa do Projeto da Gramática do Português Falado. Nessa perspectiva, de acordo com Jubran (1996a), os parênteses têm sido vistos como frases independentes inseridas em outras frases, com interrupção da estrutura sintática e sem conexão formal nitidamente estabelecida.

No tocante à semântica, no entanto, a parentetização é entendida como “um fenômeno de articulação de informações no discurso oral, e não como um fenômeno que provoca descontinuidade na progressão temática do discurso” (MINCOFF, 2008, p. 4). Essas construções, de acordo com Jubran (2006), subdividem-se em (i) estruturas com estatuto tópico, quando o conteúdo da parentética constitui uma explicação ou escl-

¹ No original: “...all of which have in common that a parenthetical is considered a linguistic entity which is linearly integrated in another linguistic structure but is unrelated to the surrounding linguistic material in one way or another, i.e. in terms of syntactic structure, semantic meaning and/or intonation.”

² No original: “an expression of which it can be argued that, while in some sense ‘hosted’ by another expression (H), P makes no contribution to the structure of H”

recimento, com relação íntima com o tema em discussão; e (ii) estruturas sem estatuto tópico, tidas como propriamente parentéticas, quando há ruptura do tópico e posterior retomada.

A autora ressalta que os parênteses podem exercer diversas funções textual-interativas, e classifica-os segundo quatro tipos: (i) parênteses focalizadores da elaboração tópica do texto, (ii) parênteses com foco no locutor, (iii) parênteses com foco no interlocutor e (iv) parênteses focalizadores do ato comunicativo. Esses quatro tipos de inserções parentéticas – que se subdividem segundo suas funções específicas – distribuem-se ao longo de um *continuum* de graus de desvio do tópico discursivo, sendo o tipo (i) aquele que menos se desvia e o tipo (iv) aquele que mais se desvia do tópico.

É importante ressaltar que, embora os trabalhos de Jubran (1996a, 1996b, 1999, 2006) e Jubran et al. (1993, 1996) façam referência a fatos prosódicos como marcas formais delimitadoras dos parênteses, essa observação não se fundamenta em análises prosódicas mais acuradas, permanecendo em um nível um tanto impressionístico. De fato, os estudos sobre as parentéticas no PB restringem-se quase exclusivamente ao campo da linguística textual, havendo pouquíssima pesquisa sobre o tema que aborde a caracterização prosódica do fenômeno.

Na sequência da apresentação do objeto de estudo realizada nesta primeira seção, este artigo conta com a seção 2, em que se faz uma caracterização prosódica das parentéticas a partir de revisão da literatura. Na seção 3, apresenta-se a fundamentação teórica e metodológica empregada na pesquisa. Em seguida, na seção 4, detalham-se os materiais e procedimentos utilizados, e na seção 5 são apresentados os resultados da análise de dados e discussão desses resultados. Por fim, na seção 6, são feitas algumas considerações finais, e propostas ideias para o aprofundamento da pesquisa sobre a prosódia das parentéticas no PB.

2 Caracterização prosódica das parentéticas: alguma literatura

Sobre a caracterização prosódica das parentéticas, destaca-se a dissertação de Mestrado de Tenani (1995), que utiliza dados de fala espontânea do Projeto da Norma Urbana Culta (NURC) e faz uma análise auditiva das marcas prosódicas de estruturas parentéticas, levando em conta parâmetros de velocidade de fala, pausa, tessitura, entonação e volume de voz, além de correlacioná-las com as marcas sintáticas características de tais estruturas. O estudo é pioneiro, na medida em que descreve um fenômeno prosódico nunca antes tratado, e o faz de um ponto de vista ainda ligado à linguística textual, utilizando-se da fonética perceptiva como ferramenta de análise.

Mais tarde, em sua tese de doutoramento (TENANI, 2002), a autora estuda os domínios prosódicos no PB, comparando dados de fala espontânea e leitura. Acerca das estruturas parentéticas, conclui que, na modalidade espontânea, não são obrigatoriamente delimitadas por pausa, enquanto, na leitura, há uma pausa no final da parentética que indica a fronteira de uma frase entoacional (IP). Os resultados de Serra (2009) para testes de percepção com dados de fala espontânea e leitura do PB corroboram essa afirmação, uma vez que as rupturas associadas a estruturas parentéticas estão entre aquelas com maior índice de reconhecimento.

Já para o português europeu (PE), os estudos das parentéticas a partir de um ponto de vista fonológico se encontram em estágio mais avançado. Frota (2000, 2014) estuda a entoação da variedade *standard* do PE, falada em Lisboa, em dados de leitura, utilizando as linhas teóricas da AM e da Hierarquia Prosódica e analisa, inclusive, a ocorrência de parentéticas. Nos trabalhos de Frota, essas estruturas são descritas, em termos prosódicos, como constituindo um IP próprio, independente do restante do enunciado, e cercado por dois tons de fronteira, um inicial e um final.

Cruz (2013), em sua tese de doutoramento, estuda a variação prosódica em duas variedades centro-meridionais do PE, abordando também as parentéticas. A autora também considera que as estruturas formam IPs independentes, com fronteiras marcadas por fenômenos segmentais, temporais e entoacionais característicos desses constituintes.

A dissertação de Mestrado de Barros (2014) se debruça especificamente sobre a questão das parentéticas, a partir de uma perspectiva teórica que conjuga a Teoria Autossegmental e Métrica e os princípios da Hierarquia Prosódica, na mesma linha dos trabalhos de Cruz (2013) e de Frota (2000, 2014). Inserido no âmbito do projeto InAPoP (*Interactive Atlas of the Prosody of Portuguese*), o trabalho analisa o fraseamento prosódico e a entoação em enunciados com construções parentéticas e tópicos em duas variedades do PE. A autora verifica uma tendência de as parentéticas formarem IPs separados e aponta, ainda, maior concentração de pausas e valores mais altos de gama de variação de F0 (ou seja, a diferença entre o valor mais alto e o valor mais baixo na curva de F0 da palavra nuclear do IP) na fronteira que precede a parentética.

3 Fundamentação teórica e metodológica

Como fundamentação teórico-metodológica, foram utilizados os pressupostos da teoria Autossegmental e Métrica da Entoação (AM) (PIERREHUMBERT, 1980; LADD, 2008; FROTA et al., 2015, entre outros), a partir da qual se fez a análise e nota-

ção entoacional dos enunciados. Recorreu-se também ao arcabouço teórico da Fonologia Prosódica, que propõe a segmentação do fluxo da fala em domínios prosódicos gramaticalmente estruturados, na linha dos trabalhos de Selkirk (1984, 1986), Nespor e Vogel (2007 [1986]), entre outros. Ademais, a metodologia inclui os procedimentos da Fonética Acústica Experimental (BARBOSA; MADUREIRA, 2015), aplicados à análise dos dados no programa computacional *Praat* (BOERSMA; WEENINK, 1992-2015).

Emprega-se, portanto, uma visão integrada, em que a entoação atua como uma das pistas que sinalizam a existência de uma estrutura prosódica (FROTA, 2000). Na abordagem adotada, pressupõe-se uma hierarquia de proeminências acentuais – por exemplo, no nível da palavra, uma sílaba só pode ser considerada “forte” relativamente à outra, que é “fraca”. Assim, os aspectos básicos da entoação, isto é, as relações de proeminência acentual que caracterizam os contornos entoacionais estão, assim como os fenômenos de junctura e estabelecimento de fronteiras, evidenciando a existência da estrutura prosódica.

3.1 Teoria Autossegmental e Métrica da Entoação

O modelo autossegmental e métrico se propõe a analisar a estrutura entoacional das línguas por meio de eventos tonais, ou seja, movimentos estabelecidos entre os diferentes níveis de altura tonal que se verificam ao longo de um enunciado. Busca determinar, a partir da relação entre esses níveis, o que tem papel fonológico nas línguas – fazendo parte, portanto, da gramática entoacional das línguas naturais – e o que consiste em variação no nível fonético, relacionando-se a um mesmo padrão frásico. Na AM, os tons são classificados em níveis de altura: alto (H, de *high*) e baixo (L, de *low*), e configuram, dessa forma, elementos contrastivos que se organizam em sequência e formam os contornos melódicos, ou contornos de F0. Deve-se salientar que, de acordo com Ladd (2008 [1996]), os níveis H e L são abstrações fonológicas, não havendo, necessariamente, correspondência direta entre os níveis de altura e valores absolutos de F0.

De acordo com o aparato metodológico da AM, atribuem-se marcas notacionais específicas aos tons relacionados às sílabas portadoras de acento lexical e aos tons relacionados a fronteiras de constituintes prosódicos. Os tons associados ao acento lexical são assinalados com um asterisco (*) após a letra H ou L. Quanto aos tons relacionados às fronteiras de constituintes prosódicos, subdividem-se em dois tipos: os chamados tons de fronteira, associados a limite de sintagma entoacional (IP), são assinalados com um símbolo de porcentagem (%); e os acentos frasais, associados a limites de constituintes menores, são indicados por um travessão (–) depois do H ou L. Outros símbolos notacionais que podem acompanhar os tons H e L são o ponto de exclamação (!), represen-

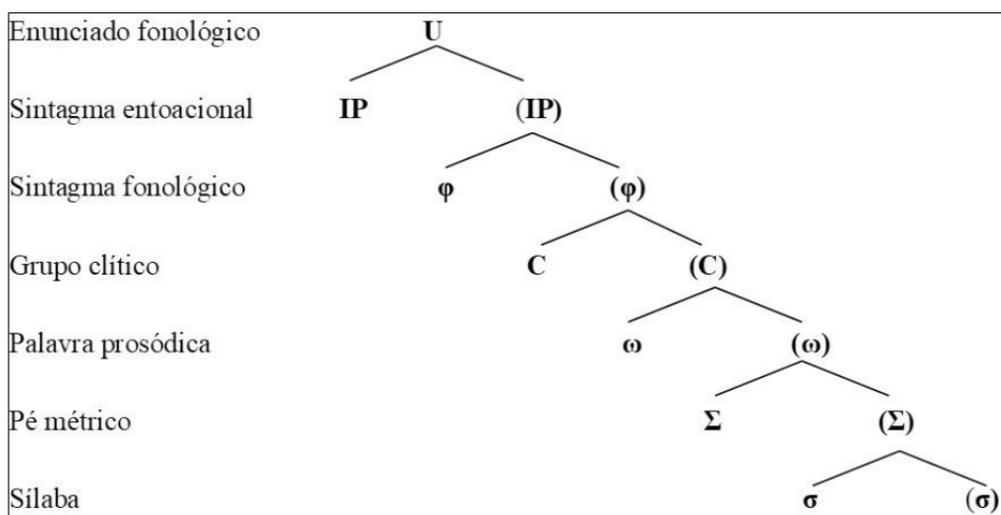
tando o processo de *downstep* – quando um tom H é relativamente mais baixo que o tom H anterior – ou o ponto de exclamação invertido ($\grave{}$), representando o processo de *upstep* – um tom H relativamente mais alto do que o tom H anterior.

A organização fonológica da entoação, segundo o modelo AM, se relaciona com a estrutura prosódica das línguas (BECKMAN; PIERREHUMBERT, 1986; LADD, 2008, entre outros). Os eventos tonais associados à fronteira de constituinte prosódico são de fundamental importância para o fraseamento prosódico – a divisão do fluxo da fala em unidades entoacionais –, uma vez que funcionam como índices de ruptura melódica, associados ou não a outras pistas como a ocorrência de pausa, o alongamento pré-fronteira, etc. (SERRA, 2009, 2016; FERNANDES-SVARTMAN et al., a sair).

3.2. Fonologia Prosódica

A Fonologia Prosódica postula que o fluxo da fala é dividido em unidades fonológicas hierarquicamente organizadas, fato que se reflete, nas línguas, em processos fonológicos tanto no nível segmental quanto no suprasegmental. A Hierarquia Prosódica é composta tradicionalmente por sete domínios, distribuídos em ordem crescente, a saber: sílaba (σ), pé métrico (Σ), palavra prosódica (ω), grupo clítico (C), sintagma fonológico (φ), sintagma entoacional (IP) e enunciado fonológico (U). Essa hierarquia costuma ser representada na forma de diagrama arbóreo (cf. Figura 1), que remete à filiação dessa teoria à Fonologia Gerativa Clássica (CHOMSKY; HALLE, 1968).

Figura 1. Hierarquia dos constituintes prosódicos



Fonte: Caldas (2018, p. 32)

Nesta pesquisa, destaca-se o papel do domínio prosódico do sintagma entoacional (IP), que compreende todas as sequências de ϕ s adjacentes numa oração matriz ou todos os ϕ s não estruturalmente ligados à oração, como é o caso das parentéticas. Do ponto de vista entoacional/prosódico, este domínio se caracteriza no PB: (i) pela presença de um acento nuclear e de um tom de fronteira no seu limite direito; (ii) pela ocorrência de um acento pré-nuclear em sua primeira palavra prosódica; e (iii) pela inserção de pausas delimitando-o e pela alta densidade de acentos tonais nas palavras prosódicas que o constituem internamente (MORAES, 2008; FROTA; VIGÁRIO, 2000; TENANI, 2002; FERNANDES, 2007; SERRA, 2009, 2016; SILVESTRE, 2012). O domínio da palavra prosódica (Pw) também recebe especial atenção, uma vez que o número de Pws à esquerda da parentética consiste em uma de nossas variáveis. O algoritmo de formação da palavra prosódica prevê que esse constituinte tem um único acento lexical (o grupo de palavra prosódica tem um único elemento proeminente). Todo elemento com acento de palavra tem de estar incluído numa palavra prosódica (VIGÁRIO, 2010).

De acordo com os princípios da Fonologia Prosódica, a estrutura fonológica está em relação com outras estruturas da gramática, principalmente com a sintática – a própria forma como o modelo assume a constituição da estrutura fonológica remete às regras de constituição sintática. A partir dessa relação, pode-se prever uma segmentação ideal do contínuo de fala em domínios prosódicos, que pode ou não se confirmar pela análise de dados concretos. É importante ressaltar, porém, que essa interação não se dá de maneira isomórfica, havendo intenso debate acerca de que tipo de informação pode ou não ser acessado no mapeamento sintaxe-fonologia (SELKIRK, 1984, 1986; NESPOR; VOGEL, 1986).

4 Materiais e procedimentos

Os dados analisados nesta pesquisa são de fala controlada, e levou-se em consideração, para a constituição da amostra, a ideia de que as estruturas parentéticas classificam-se em vários subtipos, a partir de sua função textual-interativa (JUBRAN, 2006, entre outros trabalhos). Assim, os dados analisados totalizam 14 trios de enunciados, de tamanho variável em número de sílabas e de palavras prosódicas (2Pw, 3Pw ou 4Pw), os quais possuem inserções parentéticas classificadas em dois tipos; sete dos trios de enunciados contêm parentéticas do tipo oracional, que consistem, em termos semântico-discursivos, em um comentário geral sobre a sentença (cf. exemplo 1); nos outros sete trios, as inserções são do tipo não oracional e consistem em um aposto/comentário relativo à primeira parte da oração matriz (cf. exemplo 2).

Para a elaboração dos enunciados utilizados no teste de leitura, nas fronteiras entre a parentética e as partes da oração matriz que a ladeavam, foram evitados contextos segmentais que propiciassem a aplicação de fenômenos de sândi externo. Essa decisão metodológica visava a minimizar a interação entre *texto* e *melodia* nas fronteiras dos constituintes prosódicos, na medida em que os processos de junção segmental podem atuar para a reorganização da cadeia melódica (FROTA et al., 2015), e o nosso objetivo principal era observar as estratégias de fraseamento prosódico relativamente ao tamanho do primeiro constituinte da oração matriz.

- (1) aquelas mulheres | por mais que elas neguem | precisam de ajuda
aquelas pobres mulheres | por mais que elas neguem | precisam de ajuda
aquelas pobres mulheres cansadas | por mais que elas neguem | precisam de ajuda
- (2) o irmão da vizinha | sujeito muito honesto | comprou um carro
o irmão caçula da vizinha | sujeito muito honesto | comprou um carro
o irmão caçula da vizinha fofoqueira | sujeito muito honesto | comprou um carro

Os enunciados fazem parte de um *corpus* maior de fala controlada, que serve de base para diversos estudos no âmbito do projeto de pesquisa *Fraseamento prosódico, variação e a interface prosódia-sintaxe no português do Brasil* (PPGLEV-UFRJ). O teste de leitura foi realizado por 15 falantes cariocas, todas do sexo feminino, estudantes de pós-graduação da UFRJ e com idades entre 22 e 30 anos. As informantes leram, cada uma, 127 frases, perfazendo um total de 1.860 dados produzidos, incluindo, além das estruturas com inserções parentéticas, estruturas com orações justapostas, com o sequenciador “aí”, com o marcador discursivo “né”, sentenças com /R/ em vários contextos e orações adverbiais desgarradas. Desse total, foram computados 630 dados com parentéticas, 62 dos quais foram descartados, por não apresentarem condições satisfatórias para a análise acústica/entoacional. Contamos, portanto, com 568 dados analisados, de fato, referentes às parentéticas, objeto de estudo desta pesquisa. Esses dados de parentéticas não continham qualquer tipo de sinal de pontuação que pudesse influenciar a leitura das falantes; elas puderam, entretanto, fazer a leitura silenciosa das frases antes de produzi-las.

Devido ao volume do *corpus*, cada falante realizou o teste de leitura em duas sessões, em dias diferentes. As frases eram apresentadas na tela do computador de forma randomizada (mesclando-se os tipos frásicos investigados), e a falante dispunha do tempo que

desejasse para lê-los em voz alta. O programa computacional utilizado para realização do teste foi o Psyscope X³, e a captação de som foi feita com um gravador Sony modelo PCM-D50, com capacidade de 96KHz/24bit.

Partindo de uma segmentação ideal que considera a parentética um sintagma entoacional (IP) independente (NESPOR; VOGEL, 2007 [1986]), foram observadas pistas acústicas relacionadas à presença de fronteiras prosódicas nos limites das inserções parentéticas. A análise acústica levou em conta, portanto, 1) a presença e duração de pausas silenciosas antes e depois da parentética; e 2), em cada um dos três IPs dos enunciados, a configuração da curva de F0, de acordo com os princípios da teoria AM. A partir da observação da curva de F0, realizou-se a notação entoacional dos contornos nucleares de cada um dos IPs, ou seja, a notação do movimento tonal sobre a última palavra prosódica de cada IP.

Os resultados da análise acústica e entoacional foram submetidos a tratamento estatístico através do programa R (R Core Team, 2016), para observar a distribuição dos fatores e testar a significância estatística dos resultados⁴.

5 Análise dos dados e discussão dos resultados

A ocorrência de pausas silenciosas antes e depois da parentética é, possivelmente, a pista acústica mais comumente associada a este tipo de estrutura no que concerne à prosódia (JUBRAN, 2002; SERRA, 2009). Essa ideia, que aponta para a tendência ao fraseamento das parentéticas em um IP independente do restante do enunciado, não encontra confirmação nos nossos dados: na leitura, não há, necessariamente, uma pausa silenciosa delimitando o IP da parentética, embora a ocorrência de pausa aumente conforme vai aumentando o tamanho do primeiro IP, em termos de número de palavras prosódicas. Os dados também não revelaram relação entre a ocorrência e duração de pausa ou o tipo de contorno entoacional e o tipo de parentética (oracional \times não oracional).

É interessante, nesse ponto, estabelecer contraste com os resultados de Barros (2014) para o PE: a autora afirma que, naquela variedade do português, a pausa ocorre mais frequentemente na primeira fronteira delimitadora da parentética. Nos nossos dados, os percentuais gerais de ocorrência de pausa são muito próximos na primeira e na segunda fronteiras que ladeiam a parentética, sendo a segunda fronteira (entre a parentética e o IP à direita), inclusive, ligeiramente mais demarcada por pausa do que a primeira (cf. Tabela 1).

³ Agradecemos ao Prof. Marcus Maia e ao Laboratório de Psicolinguística Experimental (LAPEX/UFRJ) pela valiosa contribuição na montagem do experimento e pelo empréstimo dos equipamentos.

⁴ Agradecemos à Juliane Venturelli, da UFMG, pelo auxílio na escolha e execução dos testes estatísticos.

Tabela 1. Percentual de ocorrência e média de duração das pausas nas duas fronteiras da parentética por número de palavras prosódicas no primeiro IP

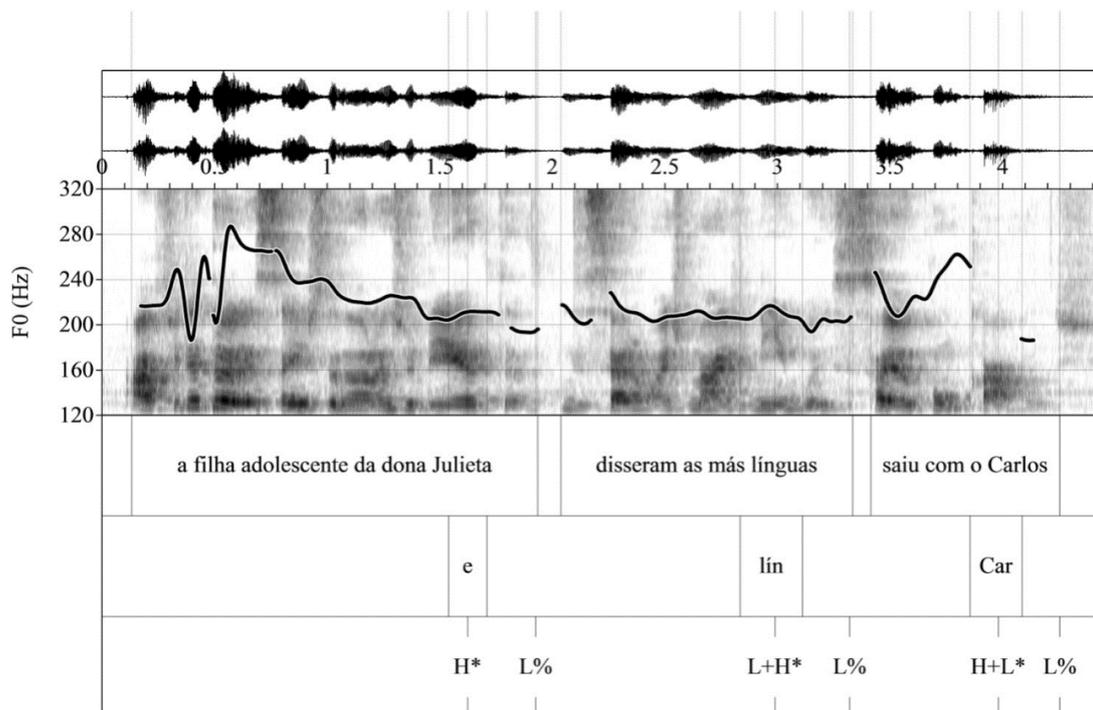
Tamanho do 1º IP	1ª fronteira		2ª fronteira	
	Dur. pausa	% oco.pausa	Dur. pausa	% oco.pausa
2pw	0,167 s	27,8%	0,164 s	36,7%
3pw	0,176 s	33,5%	0,153 s	40,8%
4pw	0,190 s	44,4%	0,166 s	40,7%
Geral	0,179 s	34,8%	0,160 s	39,4%

Fonte: As autoras (2018).

Na Figura 2, por exemplo, verifica-se um constituinte com quatro Pw no primeiro IP, em que ocorrem pausas silenciosas em ambas as fronteiras: antes e depois da parentética.

Figura 2. Enunciado 07C, informante BA

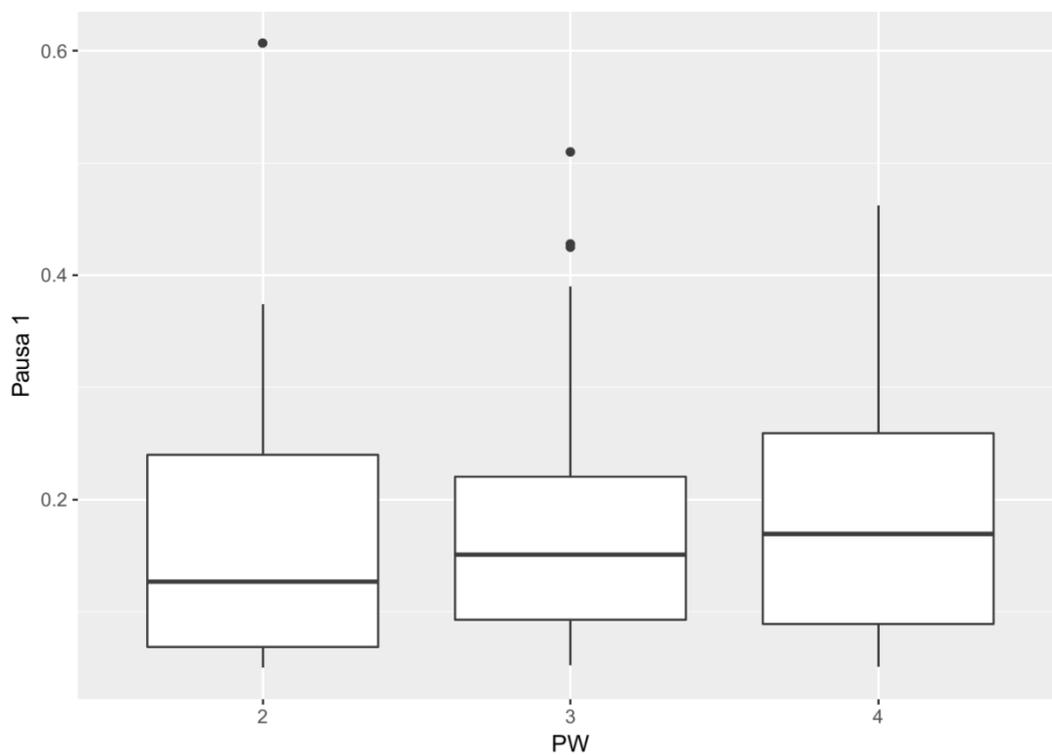
[A filha adolescente da Dona Julieta]_{IP} [disseram as más línguas]_{IP} [saiu com o Carlos]_{IP}.



Fonte: As autoras (2018).

No tocante à ocorrência e duração de pausas silenciosas, o teste estatístico revelou que o tamanho do IP à esquerda da parentética mostrou-se relevante para a demarcação de fronteiras duracionais. Na Tabela 1, é possível verificar que os valores referentes à duração e ao percentual de ocorrência de pausas na primeira fronteira crescem gradativamente conforme aumenta o número de palavras prosódicas no primeiro IP do enunciado. Em outras palavras, pode-se afirmar que, quanto maior o primeiro IP, maior a frequência e duração das pausas silenciosas na primeira fronteira.

Gráfico 1. *Boxplot* de distribuição da duração da primeira pausa pelo número de Pw no primeiro IP do enunciado



Fonte: As autoras (2018).

No Gráfico 1, acima, pode-se visualizar a distribuição das pausas da primeira fronteira: a linha que divide cada uma das caixas representa a mediana, cujo valor aumenta conforme aumenta o tamanho do primeiro IP.

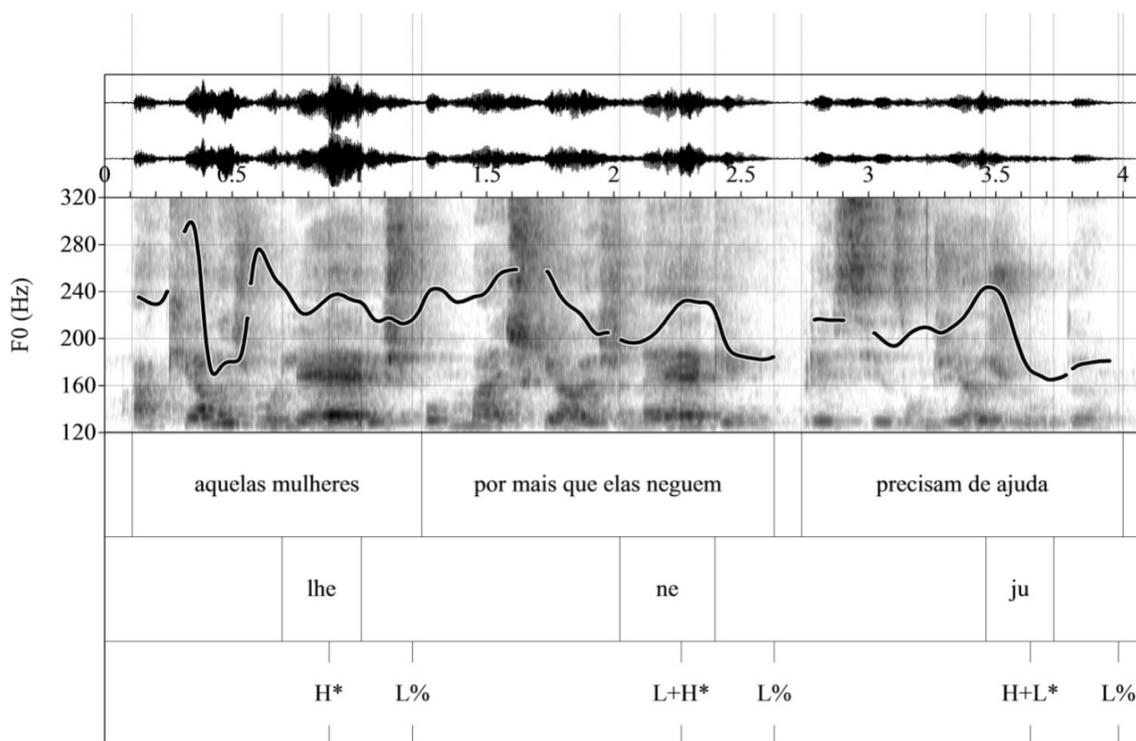
A fim de verificar se os resultados eram estatisticamente significativos, realizaram-se os testes estatísticos (i) Chi Quadrado, para testar a associação entre ocorrência de pausa e número de Pw no primeiro IP e (ii) Anova, para testar a igualdade das médias de pausa nos diferentes tamanhos de enunciado. O teste Chi Quadrado revelou evidências

de associação entre quantidade de Pw e ocorrência da primeira pausa (p-valor = 0,0110); mas não entre quantidade de Pw e ocorrência da segunda pausa (p-valor = 0,6956). Já o teste Anova não revelou diferença significativa entre as médias de duração da primeira (p-valor = 0,572) e da segunda pausa (p-valor = 0,699).

No que se refere à análise entoacional dos contornos nucleares, mais uma vez, nossos dados mostraram, para o PB, uma tendência oposta àquela verificada por Barros (2014) para o PE. A autora encontrou predomínio de tons de fronteira altos e movimentos ascendentes nos dois primeiros IPs dos enunciados. Nos nossos dados, diferentemente, predominam os tons de fronteira baixos e movimentos descendentes ou circunflexos, também nos dois primeiros IPs (cf. Figura 3). O terceiro IP do enunciado, tanto no PE como no PB, caracteriza-se quase categoricamente por um contorno descendente.

Figura 3. Enunciado 01A, informante AA

[Aqueles mulheres]_{IP} [por mais que elas neguem]_{IP} [precisam de ajuda]_{IP}.



Fonte: As autoras (2018).

Observou-se grande variedade de acentos tonais nucleares, tanto no primeiro IP quanto no IP da parentética (cf. Quadro 1 e Tabela 2). Realizou-se, no programa R, o cruzamento das variáveis (i) tipo de parentética, (ii) movimento tonal no primeiro IP e (iii) movimento tonal no segundo IP. No entanto, não foi possível verificar qualquer

relação entre o tipo de contorno nuclear realizado e o tamanho do primeiro IP ou o tipo de parentética.

Quadro 1. Contornos nucleares encontrados nos dados

Contorno	Etiqueta
H*H%	Planificado Alto
L*L%	Planificado Baixo
L*H%	Ascendente
L*+HH%	Ascendente
L+H*H%	Ascendente
L*+LH%	Ascendente
H*L%	Descendente
H*H L%	Descendente
H*+LL%	Descendente
H+L*L%	Descendente
H*+LH%	Descendente-Ascendente
H+L*H%	Descendente-Ascendente
H+L*LH%	Descendente-Ascendente
H*+LLH%	Descendente-Ascendente
L*+HL%	Circunflexo
L+H*L%	Circunflexo
L*+HHL%	Circunflexo
L+H*HL%	Circunflexo

Fonte: As autoras (2018).

Tabela 2. Distribuição percentual geral dos contornos nucleares nos três IPs dos enunciados

Contorno	Oco. 1º IP	% 1º IP	Oco. 2º IP	% 2º IP	Oco. 3º IP	% 3º IP
Ascendente	114	20,1%	129	22,8%	1	0,2%
Circunflexo	170	30%	163	28,7%	1	0,2%
Desc.-Asc.	77	13,6%	73	12,9%	1	0,2%
Descendente	176	31,1%	165	29,1%	561	98,8%
Plano Alto	5	0,9%	10	1,7%	0	0%
Plano Baixo	16	2,8%	20	3,5%	4	0,7%
N/A	9	1,5%	7	1,3%	0	0%
Total Geral	568	100%	568	100%	568	100%

Fonte: As autoras (2018).

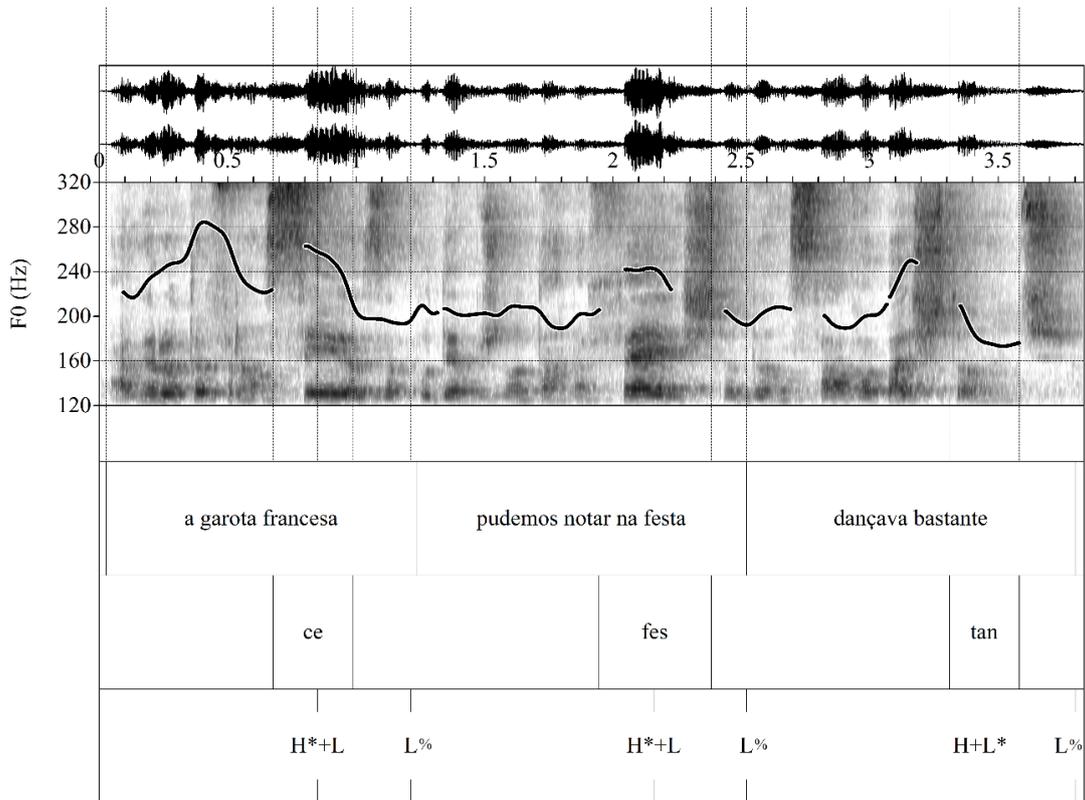
Analisando a Tabela 2, a primeira observação a se fazer é a de que, em termos de ocorrência de fronteira melódica, a parentética funciona como um IP independente, sendo delimitada por tons de fronteira tanto à sua direita (98,7%) quanto à sua esquerda (98,5%) (neste último caso, o tom de fronteira está associado ao final do IP anterior), mesmo com frequência relativamente baixa de pausas nas fronteiras. De 100% dos dados para cada contexto, somente em 1,5% (9) não ocorreu fronteira melódica entre o primeiro IP e a parentética, e somente em 1,3% dos dados (7) não ocorreu um tom de fronteira no IP da parentética.

Esses resultados demonstram 1) que nem sempre a pausa silenciosa irá demarcar as estruturas parentéticas na fala lida e 2) que o mapeamento sintaxe-prosódica é robusto, no que se refere à delimitação prosódica das parentéticas. De acordo com o algoritmo de formação do sintagma entoacional, todas as sequências de sintagmas fonológicos (φ) em uma oração raiz formam um IP e toda sequência não estruturalmente anexada à oração raiz (NESPOR; VOGEL, 2007 [1996]) deve constituir um IP próprio; é o caso de estruturas parentéticas, vocativos, *tag questions* etc. Embora a formação de IP esteja sujeita a condições de tamanho prosódico, no sentido de sintagmas longos (em número de sílabas e de palavras prosódicas) comumente serem divididos e sintagmas pequenos poderem formar um único IP com um IP adjacente, o que leva à formação de sintagmas com tamanhos equilibrados (FROTA, 2000; SERRA, 2009), em nossos dados, mesmo o menor IP das estruturas analisadas por nós, aquele que contém 2Pw, não foi reestruturado com o IP seguinte. Parece-nos que o IP parentético, portando conteúdo semântico-discursivo que se desvia do tópico discursivo mais geral da sentença, necessita sempre estar marcado prosodicamente, confirmando o que indicam os estudos sobre os parênteses na perspectiva textual-discursiva, e independentemente de sua constituição sintática interna ser oracional ou não oracional.

Em termos de configuração melódica, destacou-se, na amostra, a ocorrência dos contornos circunflexo e descendente, ao final tanto do primeiro quanto do segundo IP: em torno de 30% de cada tipo em ambos os contextos. No que se refere ao IP final, confirma-se, mais uma vez, a preferência pelo movimento melódico descendente (98,8%), típico das declarativas neutras no português (Tabela 2). A Figura 4, a seguir, ilustra a ocorrência de contornos descendentes nos dois primeiros IPs, e a Figura 5 é um exemplo de ocorrência de contornos circunflexos.

Figura 4. Enunciado 04A, informante AA

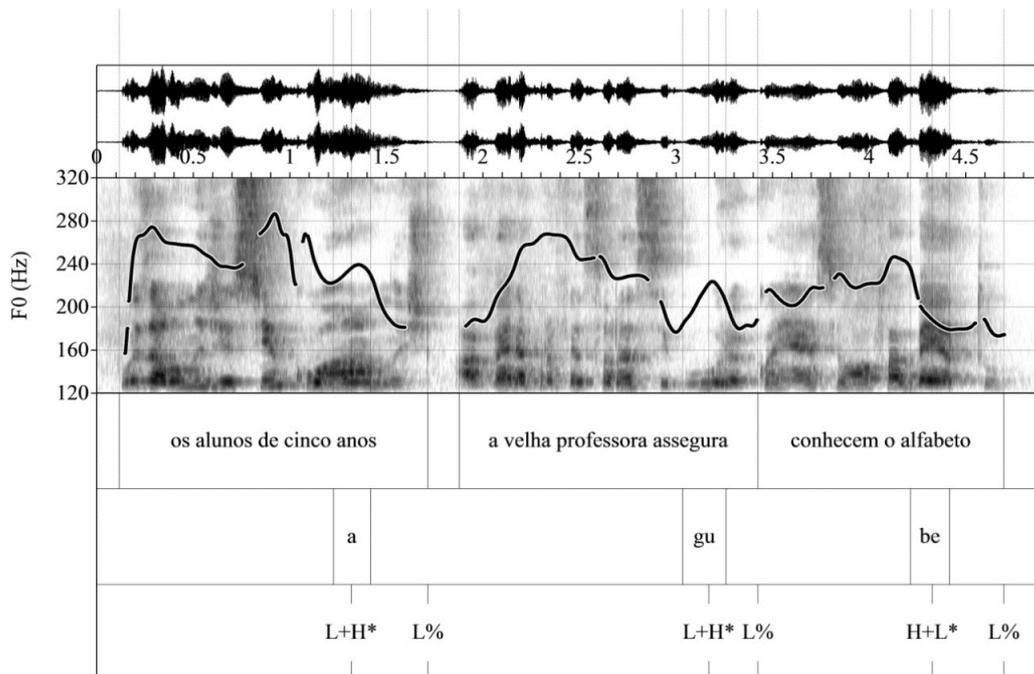
[A garota francesa]_{IP} [podemos notar na festa]_{IP} [dançava bastante]_{IP}



Fonte: As autoras (2018).

Figura 5. Enunciado 06B, informante KX

[Os alunos de cinco anos]_{IP} [a velha professora assegura]_{IP} [conhecem o alfabeto]_{IP}

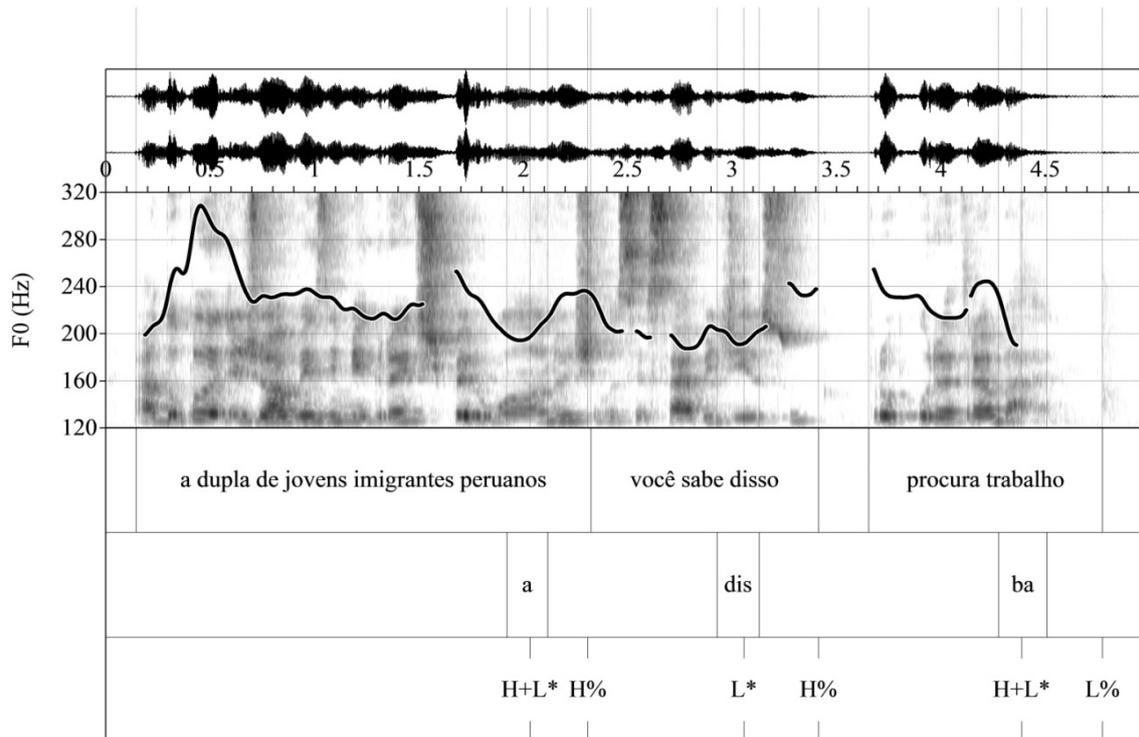


Fonte: As autoras (2018).

É válido observar, no entanto, que há tendências individuais marcadas: três das quinze informantes contrariam a tendência geral e apresentam predomínio de tons de fronteira altos e movimentos ascendentes (cf. Figura 6).

Figura 6. Enunciado 02C, informante HT

[A dupla de jovens imigrantes peruanos]_{IP} [você sabe disso]_{IP} [procura trabalho]_{IP}



Fonte: As autoras (2018).

Ao que parece, não há relação entre o tipo de parentética (*i.e.*, tipo de estrutura sintática / relação semântica entre o conteúdo da parentética e a hospedeira) e a ocorrência de tom de fronteira alto ou baixo. Na realidade, os resultados da distribuição de tons altos e baixos foram extremamente similares para as parentéticas do tipo 1 e do tipo 2 (cf. Tabela 4).

Tabela 3. Distribuição de tons de fronteira altos e baixos no segundo IP por tipo de parentética, excluídos os dados sem tom de fronteira

Tipo de parentética	Tom Alto		Tom Baixo	
	oco.	%	oco.	%
(1) Comentário	197	34,2%	212	65,8%
(2) Aposto	363	36%	348	64%
Total	560	100%	560	100%

Fonte: Elaborado pela autora.

É interessante notar que, nos dados analisados, houve maior ocorrência de contornos ascendentes e de tons de fronteira altos nos IPs aos quais não se sucede uma pausa silenciosa (cf. Tabelas 4 e 5). Sendo esses os movimentos melódicos típicos do contorno continuativo no português do Brasil (CUNHA, 2000; TENANI, 2002; SERRA, 2009, 2016; SILVESTRE, 2012), o falante parece dispensar a presença da pausa para indicar o fim (só) da unidade entoacional, que na verdade indica a continuidade do discurso.

Tabela 4. Contorno nuclear do primeiro IP *versus* ocorrência de pausa após o primeiro IP

Contorno 1o IP	Sem pausa	% sem pausa	Com pausa	% com pausa	Total geral
Ascendente *	83	22,5%	31	15,7%	92
Circunflexo	113	30,5%	57	28,9%	137
Descendente	91	24,5%	86	43,4%	142
Plano alto *	2	0,7%	2	1,3%	4
Plano baixo	9	2,4%	7	3,8%	13
Desc.-Asc. *	63	17,1%	14	6,9%	62
N/A	9	2,4%	-	-	7
Total geral	370	100%	198	100%	568

Legenda: (*) – contornos com tom de fronteira alto

Fonte: As autoras (2018).

Ainda na Tabela 4, verifica-se a maior ocorrência de contornos ascendentes, no primeiro IP, nos dados em que não ocorre pausa (22,5% contra 15,7%) e, se juntarmos os dados dos três tipos de contornos que contam com tons de fronteira altos (ascendente, plano alto e descendente-ascendente), verificamos que os tons H% totalizam 40,3% dos

dados com pausa, e apenas 23,9% dos dados sem pausa, no que se refere ao primeiro IP. O teste estatístico Chi Quadrado revelou evidências de associação entre essas variáveis, com p-valor = 0.000049.

Na Tabela 5, verifica-se o mesmo em relação ao segundo IP. Somando-se os contornos com tom H%, temos um contraste ainda maior: 50,7% de tons altos nos dados sem pausa contra 31,7% de tons altos nos dados com pausa. Como dissemos, esse resultado pode se dever ao fato de que os contornos ascendentes, no português, estão tipicamente associados a um padrão melódico continuativo, enquanto os contornos descendentes estão associados às assertivas/declarativas neutras. Novamente, o teste Chi Quadrado revelou significância estatística da associação entre as variáveis, com p-valor = 0.015319.

Tabela 5. Contorno nuclear do segundo IP *versus* ocorrência de pausa após o segundo IP

Contorno 2o IP	Sem pausa	%sem pausa	Com pausa	%com pausa	Total geral
Ascendente *	86	24,9%	44	19,4%	104
Circunflexo	91	26,4%	72	32,2%	131
Descendente	93	27,1%	72	32,2%	133
Plano alto *	6	1,8%	4	1,7%	8
Plano baixo	11	3,3%	9	3,9%	16
Desc.-Asc. *	50	14,4%	24	10,6%	59
N/A	7	2,2%	-	-	6
Total geral	344	100%	224	100%	598

Legenda: (*) – contornos com tom de fronteira alto

Fonte: As autoras (2018).

Os resultados com base na análise dos contornos melódicos confirmam a hipótese inicial de que as parentéticas constituem um sintagma entoacional (IP) independente, pelo menos na fala lida e considerando as variáveis aqui testadas. Além disso, levando em conta a pista acústica da pausa, o fraseamento prosódico mostrou-se sensível ao tamanho do primeiro IP: o total de ocorrências de pausas e suas durações aumentaram progressivamente nos dados, conforme o acréscimo do número de Pw no primeiro IP. Não foi possível, porém, observar qualquer relação entre o tamanho do primeiro IP e os tipos de acentos nucleares. Também não verificamos qualquer relação entre o tipo de estrutura interna da parentética (oracional *x* não oracional) e o seu fraseamento prosódico.

No que se refere então à relação entre a função textual-discursiva da parentética, sua característica sintática (ser um constituinte gerado fora da oração matriz) e sua realização prosódica, parece ser confirmada a intuição dos autores em geral no sentido de considerá-la

uma entidade linguística que é linearmente integrada em outra estrutura linguística, mas não é relacionada ao material linguístico circundante de uma forma ou de outra, i. e., em termos de estrutura sintática, significado semântico e/ou entoação (DEHÉ, 2014, p. 1).

Um fato importante a ressaltar é o de que a caracterização prosódica das parentéticas no PB, de acordo com nossos dados, difere daquela já descrita na literatura para o PE – tanto em relação às pausas quanto em relação à descrição tonal. Confirma-se, portanto, a tendência à diferenciação entre as gramáticas entoacionais dessas duas variedades do português, também em termos de fraseamento prosódico.

6 Considerações finais e propostas para o aprofundamento da pesquisa

Vislumbrando um aprofundamento da pesquisa sobre a prosódia das estruturas parentéticas no PB, podemos propor outras análises de dados desse mesmo tipo de estrutura: no que concerne à verificação de pistas acústicas, um estudo que leve em conta a velocidade de fala e o alongamento silábico no fraseamento das parentéticas pode trazer resultados interessantes – tais pistas, no entanto, não foram analisadas neste artigo, pois isso exigiria a elaboração de uma nova amostra de fala, com maior controle do conteúdo segmental. Também seria interessante realizar testes de percepção com ouvintes *naïves*, pois, mesmo quando não se identificam tom de fronteira ou pausa, parece haver ruptura prosódica. Por fim, em etapa posterior, seria desejável realizar a análise de dados de fala espontânea e comparação com os resultados para fala controlada.

Referências

BARBOSA, Plínio; MADUREIRA, Sandra. *Manual de fonética acústica experimental: aplicações a dados do português*. São Paulo: Cortez, 2015.

BARROS, Nádia. *Fraseamento prosódico em Português: uma análise entoacional de enunciados com parentéticas e tópicos em duas variedades do Português Europeu*. Dissertação de Mestrado em Linguística. Lisboa: Universidade de Lisboa, 2014.

BECKMAN, Mary E.; Pierrehumbert, Janet B. Intonational Structure in Japanese and English. *Phonology Yearbook III*. 1986. p. 15-70.

BOERSMA, Paul; Weenink, David. *Praat: doing phonetics by computer*. Versão 6.0.06. Disponível em: www.praat.org.

BURTON-ROBERTS, Noel. Parentheticals. In: BROWN, E. K. (Ed.). *Encyclopaedia of Language and Linguistics*. Elsevier Science, 2006. p. 179-182

CALDAS, Vitor Gabriel. *“Ai” sequenciador na fala espontânea: fraseamento prosódico e entoação no português brasileiro*. Dissertação de Mestrado. UFRJ/FL, 2018.

CHOMSKY, Noam; HALLE, Morris. *The sound pattern of English*. New York: Harper and Row, 1968.

CRUZ, Marisa. *Prosodic variation in European Portuguese: phrasing, intonation and rhythm in central-southern varieties*. Tese de doutoramento. Lisboa: UL, 2013.

DEHÉ, Nicole. *Parentheticals in spoken English: the syntax-prosody relation*. New York: Cambridge University Press, 2014.

FERNANDES, Flaviane Romani. *Ordem, focalização e preenchimento em português: sintaxe e prosódia*. Tese de Doutorado em Linguística. Campinas: IEL/UNICAMP, 2007.

FERNANDES-SVARTMAN, Flaviane R. et al. Intonational phrasing across varieties of Portuguese. In: CRUZ, M.; OLIVEIRA, P.; FROTA, S. (Ed.) *Prosodic variation (with) in languages: intonation, phrasing and segments*. Equinox Publishing. No prelo.

FROTA, Sónia. *Prosody and focus in European Portuguese*. Phonological phrasing and intonation. New York: Garland Publishing, 2000.

_____. The intonational phonology of European Portuguese. In: JUN, S. (Ed.). *Prosodic Tipology II*. Oxford: Oxford University Press, 2014.

FROTA, Sónia; VIGÁRIO, Marina. Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In: CASTRO, R. V.; BARBOSA, P. (Ed.). *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*, v.1. Coimbra: APL, 2000. p. 533-555.

FROTA, Sónia. et al. Melodia ou Texto? Estratégias de acomodação entre melodia e texto no Português. *Diadorim: revista de estudos linguísticos e literários* v. 17, n. 2. Rio de Janeiro: UFRJ, 2015. p.12-33.

FROTA, Sónia. et al. Intonational variation in Portuguese: European and Brazilian varieties. In: FROTA, S.; PRIETO, P. (Ed.). *Intonation in Romance*. Oxford: Oxford University Press, 2015, p. 235-283.

JUBRAN, Clélia Cândida A. S. et al. Organização Tópica da Conversação. 2. ed. In: ILARI, Rodolfo (Org.). *Gramática do Português Falado*. v. II. Campinas: Ed. da Unicamp, 1993. p. 69-82.

_____. Inserção: um fenômeno de descontinuidade na organização tópica. In: CASTILHO, Ataliba T. (Org.). *Gramática do Português Falado*. v. 3. Campinas: Editora da UNICAMP. 1996a, p. 61-74.

_____. Para uma descrição textual-interativa das funções da parentetização. In: KATO, Mary A. (Org.). *Gramática do português falado*. v. V. Campinas: Editora da Unicamp / Fapesp, 1996b, p. 339-354.

_____. Funções textuais-interativas dos parênteses. In: MOURA NEVES, Maria Helena de (Org.). *Gramática do português falado*. v. VII. Campinas: Editora da Unicamp, 1999, p.131-158.

_____. Parênteses: propriedades identificadoras. In: CASTILHO, Ataliba Teixeira de & BASÍLIO, Margarida (Org.). *Gramática do português falado*. V. IV. 2.ed. Campinas, Editora da Unicamp/Fapesp, 2002, p. 411-419.

_____. Parentetização. In: JUBRAN, Clélia Cândida A. S.; KOCK, Ingedore G. Villaça. *Gramática do português culto falado no Brasil*. v. I: Construção do texto falado. Campinas: Editora da Unicamp, 2006, p. 301-357.

LADD, D. Robert. *Intonational phonology*. 2. ed. Cambridge University Press, 2008.

MORAES, João Antonio. The pitch accents in Brazilian Portuguese: analysis by synthesis. *Proceedings of Speech Prosody*. Campinas: Brasil, 2008. p. 389-398.

MINCOFF, Luciane Braz Perez. *Inserções parentéticas: mecanismos de interação social*. Anais do CELSUL. UFRGS, 2008.

NESPOR, Marina; VOGEL, Irene. *Prosodic phonology*. Dordrecht: Foris, 1986.

_____. *Prosodic Phonology*. 2. ed. Berlin/New York: Mouton de Gruyter, 2007.

PIERREHUMBERT, Janet. *The phonology and phonetics of English intonation*. PhD Thesis. Massachussets: MIT, 1980.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. Disponível em: <https://www.R-project.org/>.

SELKIRK, Elisabeth. *Phonology and syntax: the relation between sound and structure*. Cambridge: MIT Press, 1984.

_____. On derived domains in sentence phonology. *Phonology* 3. MIT Press, 1986, p. 371-405.

SERRA, Carolina Ribeiro. *Realização e percepção de fronteiras prosódicas no português do Brasil: fala espontânea e leitura*. Tese de doutorado. Rio de Janeiro: UFRJ-FL, 2009.

_____. A interface prosódia-sintaxe e o fraseamento prosódico no português do Brasil. *Joss Journal of Speech Science*, v. 5, p. 47-86, 2016.

SILVESTRE, Aline Ponciano dos Santos. *A entoação regional dos enunciados assertivos nos falares das capitais brasileiras*. Dissertação de Mestrado. Rio de Janeiro: UFRJ/FL, 2012.

TENANI, Luciani Ester. *Análise prosódica das inserções parentéticas no corpus do Projeto da Gramática do Português Falado*. Dissertação de Mestrado. Campinas: Unicamp-IEL, 1995.

_____. *Domínios prosódicos no português do Brasil: implicações para a prosódia e para a aplicação de processos fonológicos*. Tese de Doutorado. Campinas: Unicamp-IEL, 2002.

VIGÁRIO, Marina. Prosodic structure between the Prosodic Word and the Phonological Phrase: recursive nodes or an independent domain? *The Linguistic Review*, v. 27, n. 4, 2010, p. 485-530.



Data de submissão: 15/03/2018

Data de aceite: 07/06/2018

MARCAÇÃO DE FOCO ESTREITO E O ACENTO SECUNDÁRIO EM INTERROGATIVAS TOTAIS NO PORTUGUÊS DO BRASIL

PRODUCTION OF NARROW FOCUS AND THE SECONDARY STRESS
IN BRAZILIAN PORTUGUESE YES/NO QUESTIONS

Manuella Carnaval | [Lattes](#) | manuellacarnaval@gmail.com
Universidade Federal do Rio de Janeiro

João Antonio de Moraes | [Lattes](#) | jamoraes3@gmail.com
Universidade Federal do Rio de Janeiro

Albert Olivier Blaise Rilliard | [Lattes](#) | albert.rilliard@limsi.fr
Universidade Federal do Rio de Janeiro

Resumo: Este artigo tem como principal objetivo investigar, em enunciados interrogativos totais, a relação entre a marcação prosódica do foco estreito e o acento secundário, no Português do Brasil, do ponto de vista da percepção e da produção. Foi elaborado um *corpus* de fala atuada, composto por 21 enunciados, para esta finalidade, de modo que fossem contempladas palavras focalizadas com variação do número de sílabas pretônicas, apresentando condições favoráveis e não favoráveis à ocorrência do acento secundário. Essas palavras apresentaram distribuição pelos seguintes contextos prosódicos: vocábulo final do enunciado, coincidindo com cabeça de I; vocábulo cabeça de sintagma fonológico (Φ), posição mais forte; vocábulo que não constitui cabeça de Φ , situando-se em posição fraca, à esquerda, podendo o sintagma fonológico estar ao início, meio ou final do enunciado. Esses enunciados foram submetidos a um teste perceptivo, com 20 juízes, que apontou para altas taxas de reconhecimento do foco interrogativo em todos os contextos analisados. Nossa análise permitiu identificar as seguintes estratégias de focalização: a associação de um pico geminado à posição do acento secundário na palavra focalizada; a compressão temporal do pico geminado, em palavras com apenas uma pretônica, portanto, que não apresentam condições favoráveis à ocorrência do acento secundário; o alongamento da sílaba tônica da palavra focalizada.

Palavras-chave: Entoação; Foco; Interrogação; Acento secundário; Fonologia prosódica.

Abstract: This paper investigates the relation, in perception and production, between prosodic marking of narrow focus and secondary stress in yes-no questions in Brazilian Portuguese. A 21-utterance *corpus* was built, presenting systematic variations in the focalized words, in terms of position of focus in the utterance, and number of pre-tonic syllables (linked to the presence of a secondary stress). The focused word appears at the following prosodic contexts: at the end of utterance, as head of the intonation phrase, or at a strong position as head of phonological phrase (Φ), or at a weak position in ϕ . The phonological phrase may be found at the beginning, middle or final part of utterance. Stimuli were evaluated by 20 listeners for the localization of the narrow focus, with high recognition scores for all the contexts analysed. Acoustic descriptions support a description of narrow focus performed with a geminate intonation peak related to a secondary stress on the focused word, and a temporal compression of the geminated peak in words presenting just one pretonic syllable, that is, without favourable conditions for secondary stress; a significant lengthening of segmental duration was observed in the stressed syllable of the focused word.

Keywords: Intonation; Focus; Yes-no questions; Secondary stress; Prosodic phonology.

1 Introdução

A interação comunicativa exige do interlocutor o entendimento de uma proposição emitida pelo falante e esta compreensão tem como “ponto de partida” uma informação conhecida previamente pelos dois participantes da situação de comunicação, falante e ouvinte, e segue em direção ao que se constitui como informação nova, noção que remonta a Weil (1844 *apud* MITTMANN, 2012, p. 134). O enunciado é, então, caracterizado como uma estrutura bipartida e os conceitos de fundo (f) e foco (F) (VALLDUVÍ, 1992, 2003; HANN, 2002; KRUIJFFKORBAYOVÁ; STEEDMAN, 2003) abordam, respectivamente, a sua composição informacional em termos de porção não informativa (ou menos informativa) e informativa:

- (1) A: Quando o porteiro recebeu a correspondência?
B: [O porteiro recebeu a correspondência]_f[**ontem**]_F.

A partir do ilustrado no exemplo (1), podemos entender a focalização como um processo em que a porção do enunciado que corresponde à informação nova é requisitada por uma pergunta anterior (MORAES, 2006). Em enunciados interrogativos totais, a informação nova, que constitui seu foco, não é mais fornecida pelo falante, como na asserção, mas sim por ele solicitada (MORAES; CARNAVAL; COELHO, 2015), como podemos verificar na troca abaixo:

- (2) A: O porteiro recebeu a correspondência.
B: [O porteiro recebeu a correspondência]_f [ontem]_f?

Assim, o foco recai sobre uma parte da pergunta.

Com relação à extensão do termo focalizado, é consensual, na literatura especializada, a denominação de dois tipos de foco: foco amplo e foco estreito (FROTA, 2000; AVESANI; VAYRA, 2003; SITYAEV; HOUSE, 2003; LADD, 2008; MORAES; CARNAVAL; COELHO, 2015; CARNAVAL, 2017). Essa categorização foi definida em Moraes, Carnaval e Coelho (2015) como uma “dimensão contínua”, construída a partir do aumento gradual da extensão do constituinte focalizado, o que será detalhado na seção seguinte. Nesse trabalho, os autores, ao abordarem a marcação prosódica do foco em enunciados interrogativos totais, puderam observar que, nos casos de foco estreito (mais especificamente, recaindo sobre palavra fonológica), uma subida melódica suplementar com uma distância mínima de duas sílabas da tônica foi utilizada como estratégia pelos informantes do *corpus* para destacar a palavra em foco de alguns dos enunciados analisados. Essa dupla modulação melódica sobre o vocábulo sob foco estreito foi, então, associada à posição do acento secundário da palavra focalizada, tendo em vista que o acento secundário ocorre dentro do limite da palavra fonológica (LEE, 2002; MORAES, 2003a), desde que apresente a condição mínima de duas sílabas pretônicas para ser manifestado.

Sendo este um achado que nos surpreendeu na análise, procura-se no presente estudo aprofundar, com um *corpus* expressamente elaborado para esse fim, a investigação da marcação do foco estreito a partir da associação de um pico geminado à posição do acento secundário das palavras focalizadas em enunciados interrogativos totais. Nosso especial interesse nesse tipo de enunciado se deve ao reduzido número de trabalhos que investigam a realização prosódica do foco em enunciados interrogativos, bem como por ter sido verificado em trabalho anterior (MORAES; CARNAVAL; COELHO, 2015), sobre o foco prosódico em interrogativas totais, a relação entre a produção focal e o acento secundário.

Assim, os objetivos definidos neste estudo são:

- Averiguar o peso / a eficácia da manifestação do acento secundário na marcação do foco estreito.
- Verificar as alternativas possíveis para a marcação do foco estreito quando a realização do acento secundário é inibida (menos de 2 sílabas pretônicas).

2 O acento secundário no PB

Em relação ao Português do Brasil, o acento secundário tem sido tratado em diversos trabalhos, seja de uma perspectiva fonológica, como em Bisol (1992, 1994), Collischonn (1993, 1994), Lee (2002), Sandalo et al. (2006), Fernandes-Svartman et al. (2009); seja enfatizando sua manifestação fonética, como em Arantes (2010), Arantes e Barbosa (2002, 2008), Fernandes-Svartman et al. (2012), Moraes (2003a, 2003b), Rossi (1998).

Para Lee (2002), o acento secundário difere do acento primário por ser insensível à quantidade de peso silábico e à categoria lexical, sendo puramente fonológico. Assim, o acento primário tem sua aplicação no nível lexical, enquanto o acento secundário no PB possui caráter iterativo, ocorrendo regularmente a cada 2 sílabas, da direita para a esquerda, a partir do acento primário, sempre dentro do domínio da palavra fonológica (ω), como pode ser verificado nos exemplos abaixo, em que a sílaba em negrito apresenta o acento secundário e as sílabas entre parênteses o acento primário:

- a. **Deses**(pe)ro
- b. **Desespera**(dor)¹
- c. **Desespe**(ra)do ou **Desespe**(ra)do

No último exemplo (c), podem ser observados dois padrões de atribuição do acento secundário. Lee (2002) e Moraes (2003a) destacam a possibilidade de alternância de padrões quando o número de pretônicas é ímpar, ocorrendo o efeito dactílico ou pé troqueado no início da palavra, com o deslocamento do acento secundário para a primeira sílaba.

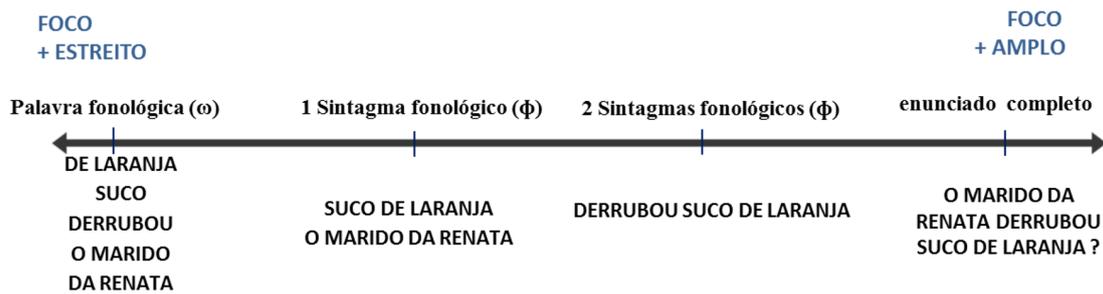
Moraes (2003, p. 148), a partir de uma abordagem acústica do fenômeno, define o acento secundário como “a proeminência que vai incidir sobre determinadas sílabas de palavra, marcando assim as fronteiras do pé”, sendo, portanto, a marca formal do pé métrico. Para o autor, o acento secundário é um fenômeno fonético-fonológico no PB, que se manifesta nas palavras com mais de uma sílaba pretônica. Dentre os parâmetros analisados, o estudo destaca a atuação da F0 na marcação acústica do acento secundário juntamente com a da duração. Esse dado direciona em nossa análise a observância de tais fatores para avaliar a relevância do acento secundário na marcação do foco estreito em enunciados interrogativos totais no PB.

¹ Na realidade, o padrão (b), com dois acentos secundários, previsto em estudos de cunho fonológico, como Lee (2002), costuma se realizar, em contextos não enfáticos, com um dos acentos secundários previstos apenas.

3 O pico geminado como correlato do foco interrogativo

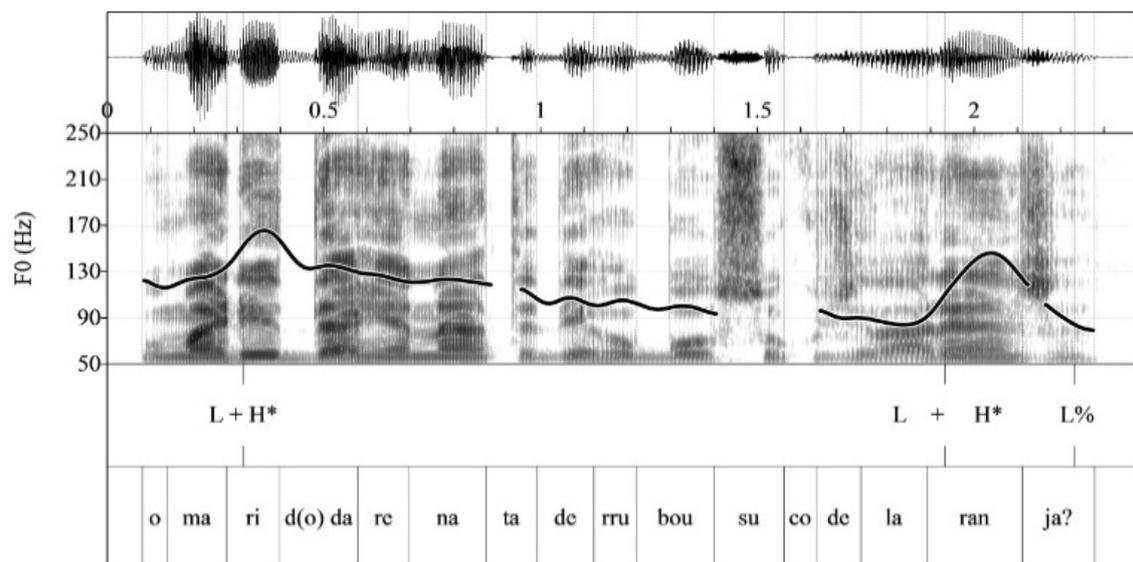
Em Moraes, Carnaval e Coelho (2015), procurou-se averiguar a marcação do foco em enunciados interrogativos totais a partir de diferentes extensões, indo do domínio mais estreito (palavra fonológica) para o mais amplo (enunciado completo), na mesma frase: “O marido da Renata derrubou suco de laranja?”, como esquematizado na Figura 1.

Figura 1. Contínuo de domínios prosódicos que podem estar sob foco, indo do foco mais estreito para o mais amplo, no enunciado “O marido da Renata derrubou suco de laranja”.



Verificou-se que o foco amplo² é caracterizado por uma dupla subida melódica, nas posições pré-nuclear, primeira sílaba tônica do enunciado, e nuclear, última sílaba tônica do enunciado, como indicado na Figura 2.

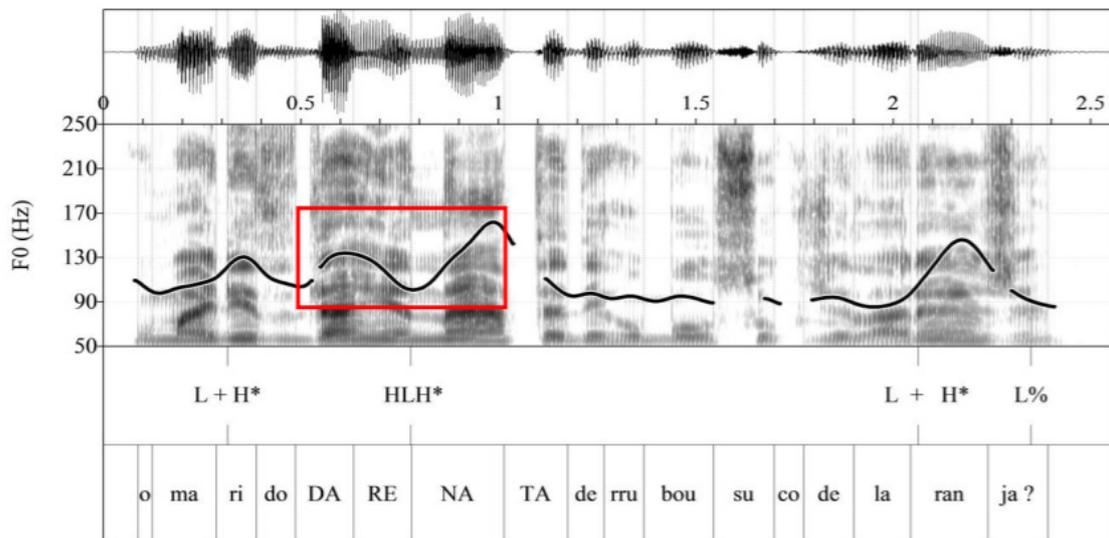
Figura 2. Contorno melódico da frase interrogativa *O marido da Renata derrubou suco de laranja?* dita com foco amplo, pelo informante C. (MORAES; CARNAVAL; COELHO, 2015).



² Entende-se por foco amplo a ausência de foco sobre um constituinte específico.

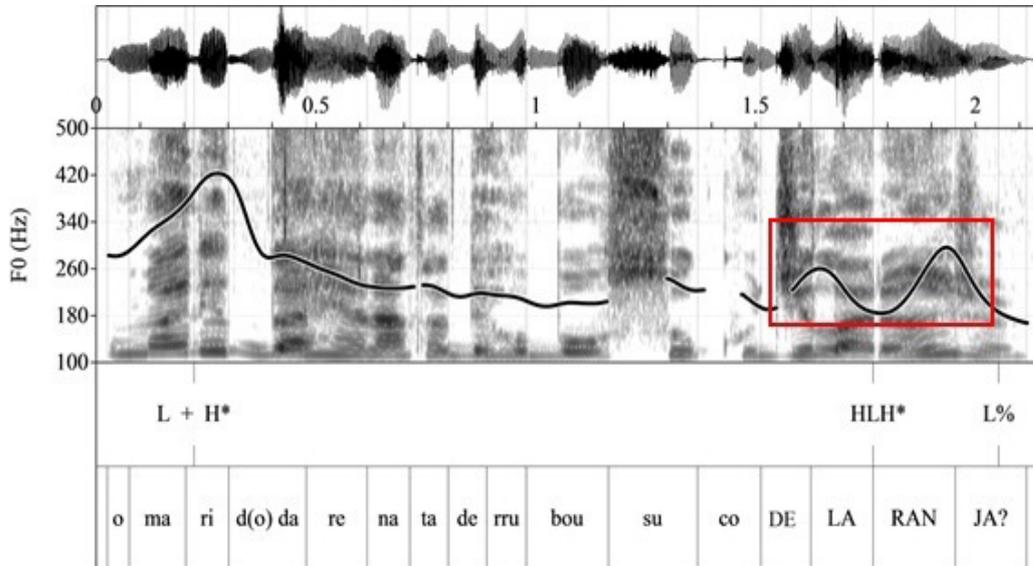
Com relação ao foco estreito, quando o foco recai sobre palavra fonológica, este pode ser marcado pelo que foi, então, por nós nomeado “pico geminado”, associado à posição do acento secundário sobre a palavra focalizada, ao qual foi atribuído o tom complexo HLH*, como pode ser observado na Figura 3. Assim, além das inflexões ascendentes no pré-núcleo e no núcleo, há, ainda, uma terceira proeminência melódica, sobre a palavra focalizada, caracterizada pelo pico geminado, cuja realização fonética ocorre em associação à posição de um possível acento secundário, em sílaba pretônica não adjacente à tônica.

Figura 3. Contorno melódico da frase *O marido DA RENATA derrubou suco de laranja?*, dita com foco sobre *da Renata*, pelo informante A; presença do pico geminado assinalada pelo retângulo em vermelho (MORAES; CARNAVAL; COELHO, 2015).



Quando a palavra focalizada ocupa a posição final do sintagma entonacional, (I) (Figura 4), a presença do pico geminado na posição final torna-se especialmente relevante para distinguir foco amplo de foco estreito, uma vez que já não há a possibilidade de uma subida melódica extra (a terceira) assinalar o vocábulo em foco, como ocorre nos demais casos, justamente pelo fato de o constituinte focalizado se situar na posição final do enunciado, que já deve receber uma subida por *default*. Assim, se entre esses dois tipos de foco é percebida uma diferença, esta será devida à presença de algum correlato especial, como o pico geminado associado ao acento secundário e/ou um comportamento diferenciado da duração, por exemplo.

Figura 4. Contorno melódico da frase *O marido da Renata derrubou suco DE LARANJA?*, dita com foco sobre *laranja*, pelo informante A; presença do pico geminado assinalada pelo retângulo em vermelho (MORAES; CARNAVAL; COELHO, 2015).



Partindo de nossas indagações específicas sobre a importância do acento secundário na marcação prosódica do foco estreito em interrogativas totais no PB, as hipóteses que pretendemos averiguar são:

Possui o acento secundário um papel legítimo na marcação do foco estreito, sendo caracterizado fonologicamente por um pico geminado no contorno melódico sobre o constituinte focalizado?

Quando sua manifestação é inibida (menos de duas sílabas pretônicas), a distinção entre foco estreito e amplo nas interrogações é neutralizada? Ou, ao contrário, o foco estreito pode ser compensatoriamente marcado pela manifestação de outros parâmetros, como o aumento da duração sobre o foco ou, até mesmo, pelo deslocamento do tom alto do acento secundário para a primeira pretônica, na eventualidade de haver uma pretônica, caracterizando um ajuste melódico ao material textual (FROTA et al., 2015)?

4 Metodologia

4.1 O corpus

Foi elaborado um *corpus* de fala não espontânea, controlada, composto pelas seguintes frases:

1. {[O namorado da Luma]ϕ [planejava beber]ϕ [refrigerante de uva]ϕ ?} I
 ω ω ω ω ω ω
2. {[O primo da Manuela]ϕ [vai derrubar]ϕ [suco de tangerina]ϕ ?} I
 ω ω ω ω ω ω
3. {[Festa cheia]ϕ [costuma dar]ϕ [refresco quente]ϕ ?} I
 ω ω ω ω ω ω

Atentando para os domínios hierárquicos prosódicos constitutivos do enunciado (NESPOR;VOGEL, 1986), procurou-se estabelecer uma composição simétrica em cada estrutura textual, de modo que consistissem em um sintagma entonacional (I) formado por três sintagmas fonológicos (ϕ). Estes, por sua vez, são compostos, cada um, por duas palavras fonológicas (ω), domínio do acento secundário (LEE, 2002), uma ocupando a posição forte, cabeça de ϕ, e outra a posição fraca, à esquerda. Pronunciadas com foco amplo (toda a frase) e foco estreito (sobre cada uma das palavras fonológicas que integram as frases), foram gerados 21 enunciados, gravados por 2 informantes (masculino e feminino, falantes da variedade do Rio de Janeiro) perfazendo um total de 42 enunciados. Esses 21 enunciados são apresentados no Quadro 1.

Quadro 1. Estímulos sonoros produzidos por cada informante a partir de cada frase, com foco amplo na primeira linha e foco estreito (em negrito) nas demais.

Frase 1	Frase 2	Frase 3
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva?	O primo da Manuela vai derrubar suco de tangerina?	Festa cheia costuma dar refresco quente?
O namorado da Luma planejava beber refrigerante de uva ?	O primo da Manuela vai derrubar suco de tangerina ?	Festa cheia costuma dar refresco quente ?

Para a obtenção destes enunciados interrogativos ditos com diferentes focos, foram criados pequenos contextos.³ Após a leitura desses contextos, os informantes produziram os diferentes enunciados apresentados acima. No *corpus*, opusemos as palavras dos enunciados a partir da quantidade de sílabas pretônicas, em cada contexto de análise, como esquematizado no Quadro 2.

Quadro 2. Palavras que compõem nosso *corpus* apresentadas segundo seus contextos estruturais e o número de pretônicas apresentado por cada uma.

Nº de pretônicas		Contextos		
		2 ou mais pretônicas	1 pretônica	Sem pretônica
Vocábulo cabeça do sintagma entonacional (I)		de tangerina 3 2 1 T	de uva 1 T	quente T
Posição forte (cabeça de ϕ)		derrubar 2 1 T da Manuela 3 2 1 T	beber 1 T da Luma 1 T	dar T cheia T
Posição fraca (à esquerda de ϕ)	ϕ final	refrigerante 3 2 1 T	refresco 1 T	suco T
	ϕ medial	planejava 2 1 T	costuma 1 T	vai T
	ϕ inicial	O namorado 3 2 1 T	O primo 1 T	Festa T

Objetivando dar continuidade à análise realizada em Moraes, Carnaval e Coelho (2015), os contextos de análise dos padrões melódicos com foco estreito foram mantidos, a saber: vocábulo final do enunciado, coincidindo com cabeça de I; vocábulo cabeça de sintagma fonológico (ϕ), posição mais forte; vocábulo que não constitui cabeça de ϕ , situando-se em posição mais fraca, à esquerda. No último contexto, consideramos, ainda, três possibilidades a partir da posição do sintagma fonológico (ϕ), ao início, meio ou final do enunciado.

O controle do número de sílabas pretônicas no domínio da palavra fonológica (ω) pretendeu propiciar, com a condição mínima de duas sílabas pretônicas, ou inibir, sem respeitar esta restrição, a manifestação do acento secundário. No segundo caso, foram consideradas duas possibilidades: (i) palavra com uma única sílaba pretônica, para averiguar o possível deslocamento do tom alto do acento secundário para a pretônica existente, e (ii) palavra sem sílabas pretônicas, onde o citado deslocamento não seria uma opção.

³ Contextos similares a esses podem ser vistos em Moraes, Carnaval e Coelho (2015, p. 173-174).

4.2 Teste de percepção

Os 42 enunciados do *corpus* foram utilizados na aplicação de um teste perceptivo de escolha forçada, realizado com 20 juízes, para validar auditivamente a marcação do foco estreito e avaliar a qualidade dos dados obtidos, compensando-se assim o tamanho relativamente reduzido do *corpus*. Esse teste foi dividido em 3 etapas, relativas aos enunciados produzidos a partir de cada uma das 3 frases. Em cada etapa, os enunciados foram dispostos de forma aleatória para os voluntários, que, após ouvi-los, deveriam selecionar dentre as sete opções disponibilizadas aquela que seria o foco da pergunta. Além das opções de foco estreito, sobre cada uma das seis palavras fonológicas que compõem os enunciados, o ouvinte poderia, ainda, selecionar a opção “sem foco”, relativa ao foco amplo, quando não identificasse foco sobre um constituinte específico.

4.3 Análise acústica

Nossa análise contempla a descrição fonética dos padrões de focalização em questão, além da atribuição de uma notação fonológica em pontos relevantes dos contornos melódicos analisados, a partir do modelo AM/ToBI (JUN, 2005, 2014), com alguma adaptação⁴. Os padrões prosódicos foram obtidos com o programa Praat de análise acústica, com o qual foram obtidas também as medidas de duração aqui consideradas.

4.3.1 Padrões melódicos

Além das figuras que ilustram visualmente o comportamento melódico da focalização, procurou-se dar uma descrição qualitativa da evolução da F0 ao longo dos enunciados, procurando, assim, apontar as estratégias e os padrões de focalização encontrados nas distintas posições em que o foco figurava no enunciado.

4.3.2 Duração

Com o programa de análise acústica Praat, os enunciados de nosso *corpus* foram segmentados por sílaba, cuja duração, medida em segundos, foi descrita aqui de maneira a contemplar as partes dos enunciados que têm maior ou menor alongamento de sílaba, sob efeito da focalização das palavras e de outros fatores que poderiam ter influência sobre o aumento de duração. As três frases que compõem nosso *corpus* foram construídas de maneira a apresentar características equilibradas do ponto de vista fonológico e sintático (cf. Seção 4.1), mas apresentam, inevitavelmente, variações na natureza dos fonemas que compõem cada sílaba. Essa característica pode, assim, interferir na duração intrínseca

⁴ Especialmente a atribuição do tom complexo HLH* ao pico geminado.

das sílabas (BARBOSA, 2007). Devido a essas considerações, o alongamento das sílabas de cada frase com foco estreito foi medido em relação à mesma sílaba na frase equivalente com foco amplo, isto é, na ausência de foco sob uma palavra específica, seguindo a equação em 1.

$$(4) \quad \delta = \log_{10} \left(\frac{d_e}{d_a} \right)$$

Nesta equação, o alongamento é medido como o logaritmo de base 10 da razão entre a duração da sílaba da frase com foco estreito (d_e) e a duração da sílaba correspondente na frase com foco amplo (ver BELIÃO, 2016, para uma abordagem comparável do alongamento, por exemplo). A utilização desse logaritmo garante que a razão seja simétrica. Segundo esse cálculo, para o caso de não haver alongamento, com a medida igual a 0, a duração da sílaba nas duas frases deve ser idêntica; uma medida de 0,3 corresponderia aproximadamente a uma duração duas vezes maior da mesma sílaba com foco estreito em relação à sílaba com foco amplo (d_a) (respectivamente, uma medida de -0,3 indicaria uma duração para o foco estreito que corresponderia aproximadamente à metade da duração do foco amplo). A medida foi aferida sobre cada sílaba das frases, seja uma sílaba tônica, postônica ou pretônica. A posição da palavra fonológica (ω) que contém a sílaba analisada, dentro da frase, constitui outro fator de variação do alongamento, sendo identificada de 1 (primeira palavra fonológica) até 6 (última palavra fonológica das frases).

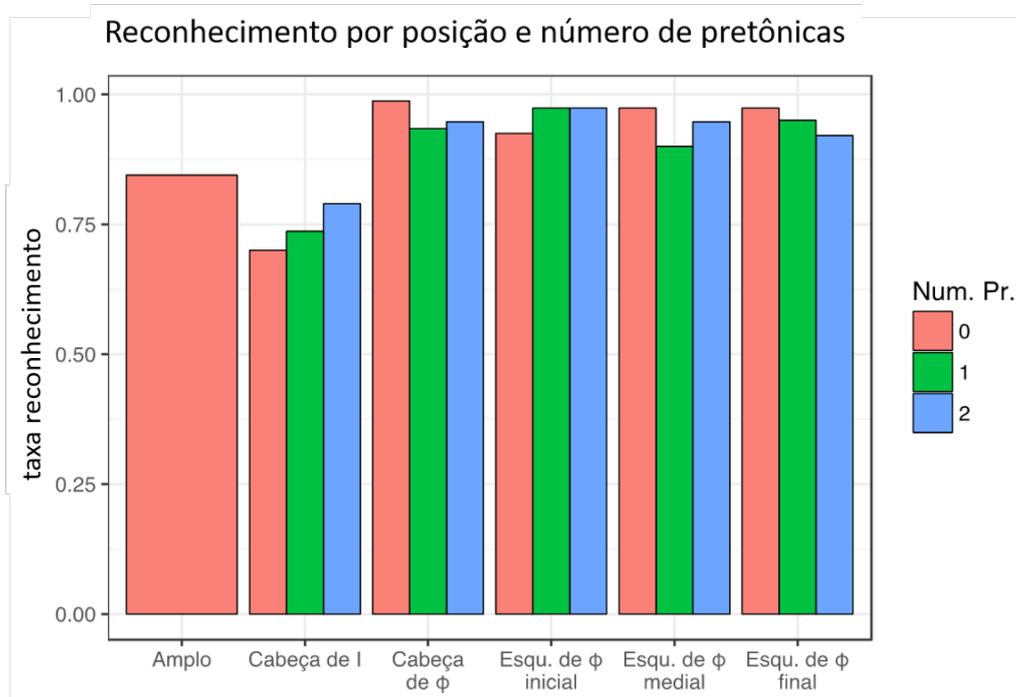
5 Resultados

5.1 Teste perceptivo

A Figura 5 ilustra os percentuais de reconhecimento do foco amplo e do foco estreito em cada contexto de ocorrência por quantidade de sílabas pretônicas apresentada. Podemos observar que todos os índices apresentados na Figura 5 são muito superiores à taxa de acerto ao acaso (14% - pois os ouvintes deviam optar por uma resposta dentre 7 disponíveis).

Para que uma análise estatística pudesse ser realizada sobre os resultados do teste perceptivo aplicado, as respostas dos juízes foram codificadas em termos de acerto e erro. Desse modo, a taxa média de acerto referente à categorização segundo a “posição” do foco (agrupando as respostas referentes aos 3 enunciados que compõem nosso *corpus*) foi utilizada como variável dependente em uma regressão logística, com os seguintes fatores: posição do foco no enunciado (6 níveis: foco amplo, cabeça de I, cabeça de ϕ , esquerda de ϕ inicial, esquerda de ϕ medial, esquerda de ϕ final); número de sílabas pretônicas (3 níveis: sem pretônica, uma pretônica, duas ou mais pretônicas); interação entre os dois fatores precedentes (ver Figura 5 para mais detalhes).

Figura 5. Taxa de reconhecimento dos constituintes focalizados; no eixo vertical, a porcentagem de identificações corretas, no horizontal, os contextos analisados apresentados de acordo com o número de pretônicas da palavra.



Um processo de simplificação de modelo (CRAWLEY, 2013), para a exclusão de fatores não significativos, mostrou que a interação ($LRT^5 = 8.0$, d.f. = 8, $p = 0.43$) e o número de sílabas pretônicas ($LRT = 0.7$, d.f. = 2, $p = 0.70$) não apresentam um efeito significativo sobre a taxa de acerto. Na verdade, o modelo ótimo explica a variação da taxa de categorização correta do foco em função de sua posição nos enunciados ($LRT = 48.2$, d.f. = 5, $p < 0.001$). Este fator apresentou, na simplificação do modelo, o agrupamento de níveis similares (cf. Figura 5). As posições foco ampla e foco sobre cabeça de I foram agrupados em um novo nível, constituindo um modelo equivalente ao precedente, com uma diferença significativa entre eles ($z = 6.5$, $p < 0.001$). Valida-se estatisticamente, então, que o reconhecimento do foco é significativamente melhor quando este ocorre sobre um elemento não final do enunciado.

Os únicos contextos que apresentam taxas inferiores a 90% são o enunciado de foco amplo e aquele cuja localização da palavra focalizada corresponde à cabeça do sintagma entonacional (I), que foram mais comumente confundidos entre si, como pode ser observado na matriz de confusão abaixo.

⁵ Os efeitos relatados aqui são baseados em *Likelihood Ratio Test (LRT)*.

Tabela 1. Matriz de confusão criada a partir dos resultados do teste perceptivo, em que se apresentam as produções dos enunciados com focalização em diferentes posições (eixo vertical) e a avaliação realizada pelos juízes no teste de percepção (eixo horizontal); na linha diagonal, o número de votos em que a intenção do falante foi bem reconhecida pelos ouvintes.

Tipo de foco	Resposta					
	Amplo	Cabeça de I	Cabeça de ϕ	Esqu. ϕ inic	Esqu. ϕ med	Esqu. ϕ final
Amplo	84	8	5	3	0	0
Cabeça de I	20	74	2	3	0	1
Cabeça de ϕ	1	0	96	3	0	0
Esqu. ϕ inic	2	0	3	96	0	0
Esqu. ϕ med	3	1	2	0	94	0
Esqu. ϕ final	3	0	1	2	0	95

No entanto, seu reconhecimento é, ainda assim, acima de 60% em todas as possíveis variações do número de pretônicas. Os índices de reconhecimento do constituinte focalizado, superiores a 60%, em todos os contextos, e a 90%, na maioria dos casos, independentemente do número de sílabas pretônicas apresentado, indicam que a marcação prosódica do foco estreito apresenta alta produtividade em interrogativas totais no PB, sendo a palavra fonológica o seu domínio por excelência.

O fato de até mesmo as palavras que não apresentam a condição requerida para a realização do acento secundário (mínimo de duas pretônicas) serem fortemente reconhecidas é um indicativo de que a língua estabelece um sistema de compensação na manifestação das estratégias de focalização. Na subseção seguinte, nossa análise procura definir que estratégias estão atuando na marcação do foco estreito em interrogativas totais nos referidos contextos, observando o número de pretônicas apresentado pelos vocábulos fonológicos para, assim, investigarmos também a participação do acento secundário nesse fenômeno.

5.2 Análise fonética e fonológica da F0

5.2.1 Focalização sobre cabeça do sintagma entonacional (I)

Quando o vocábulo focalizado ocupa a posição final do enunciado, coincidente com a posição de cabeça do sintagma fonológico (I), podemos dizer que há uma situação potencialmente ambígua, visto que se trata da posição que deverá manifestar também a natureza

modal do enunciado interrogativo, marcada pelo contorno nuclear. Assim, a ocorrência de uma proeminência sobre esse ponto não seria suficiente para assinalar o foco estreito, como foi apontado em Moraes, Carnaval e Coelho (2015), sendo necessário um correlato especial para diferenciá-lo do foco amplo, como, por exemplo, um pico geminado de tom complexo (HLH*), cujo primeiro tom alto está associado à posição do acento secundário sobre pretônica não adjacente à tônica.

De fato, em nosso *corpus*, nesse contexto, esta modulação característica também foi observada: no foco estreito em cabeça de (I), para palavra com duas ou mais sílabas pretônicas, podemos ressaltar que os dois informantes apresentam o pico geminado, de tom complexo HLH* em posição nuclear, assinalando sua distinção em relação ao foco amplo. A proeminência inicial, *default*, é mantida, apresentando aqui uma variação, com o deslocamento do tom alto para a sílaba postônica, L +> H*, no foco amplo (Figura 6) e L + H* no foco estreito (Figura 7).

Figura 6. Forma de onda e curva de F0 suavizada (bandwidth 10Hz) do enunciado *O primo da Manuela vai derrubar suco de tangerina?* pronunciado com foco amplo pelo informante masculino. No quadro à direita, nessa e nas demais figuras, o percentual de reconhecimento correto quanto à localização de foco, obtido pelo enunciado no teste perceptivo.

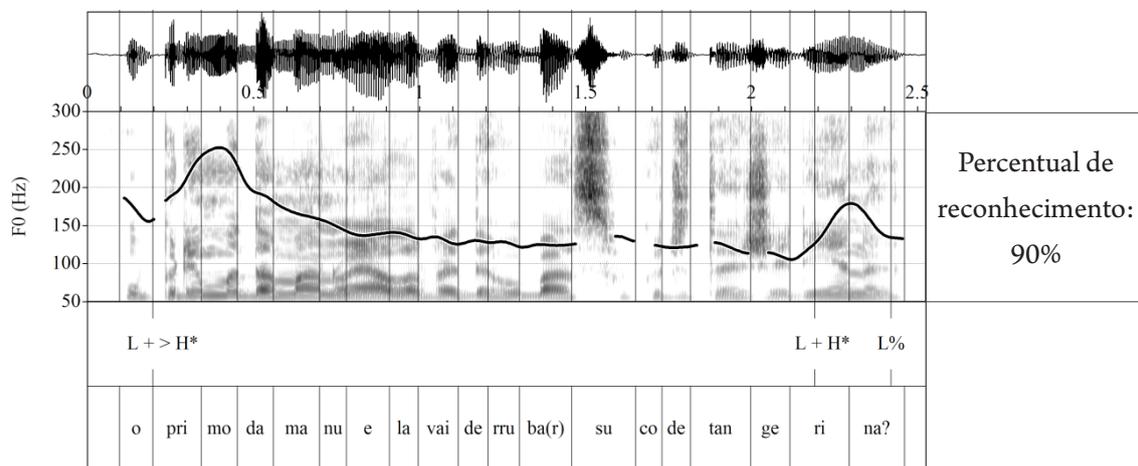
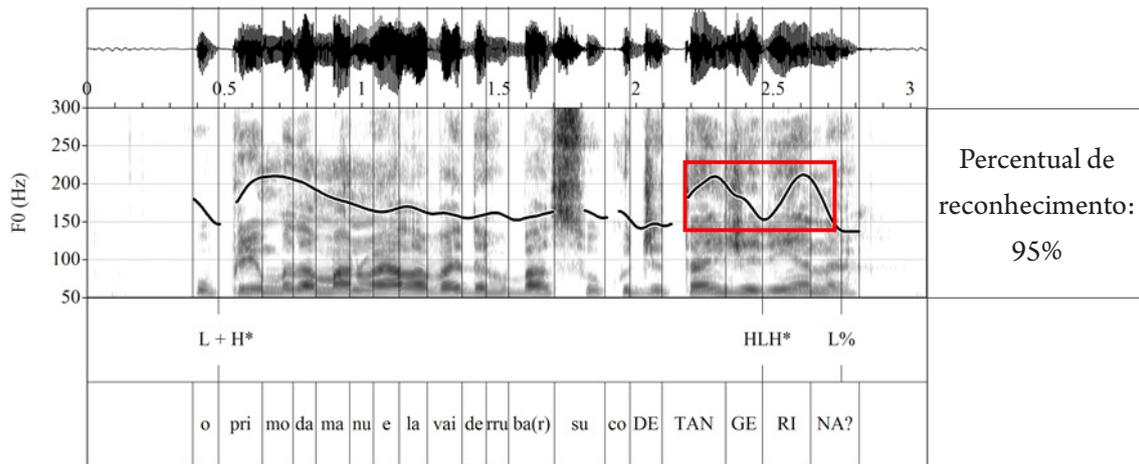


Figura 7. Forma de onda e curva de F0 suavizada (bandwidth 10Hz) do enunciado *O primo da Manuela vai derrubar suco DE TANGERINA?* pronunciado com foco sobre DE TANGERINA pelo informante masculino; presença do pico geminado assinalada pelo retângulo em vermelho.



Para averiguar como o foco estreito é assinalado quando a condição canônica de ao menos duas pretônicas para a manifestação do acento secundário não é atendida, analisamos os enunciados em que (i) a palavra focalizada apresenta apenas uma pretônica e (ii) a palavra focalizada não apresenta pretônica.

Em (i), na focalização sobre DE UVA (Figura 9), pode-se observar ainda assim a presença do pico geminado, embora comprimido temporalmente, com uma primeira subida sobre a pretônica (de), seguida de uma ligeira depressão melódica no início da sílaba seguinte, a tônica [u], para então apresentar seu segundo pico na parte final dessa sílaba, mantendo, assim, o tom complexo HLH*, caracterizando o que se poderia considerar um pico geminado (ou um acento secundário) comprimido, o que se torna evidente quando comparado ao contorno do foco amplo (Figura 8).

Figura 8. Forma de onda e curva de F0 suavizada (bandwidth 10 Hz) do enunciado *O namorado da Luma planejava beber refrigerante de uva?* pronunciado com foco amplo pelo informante masculino.

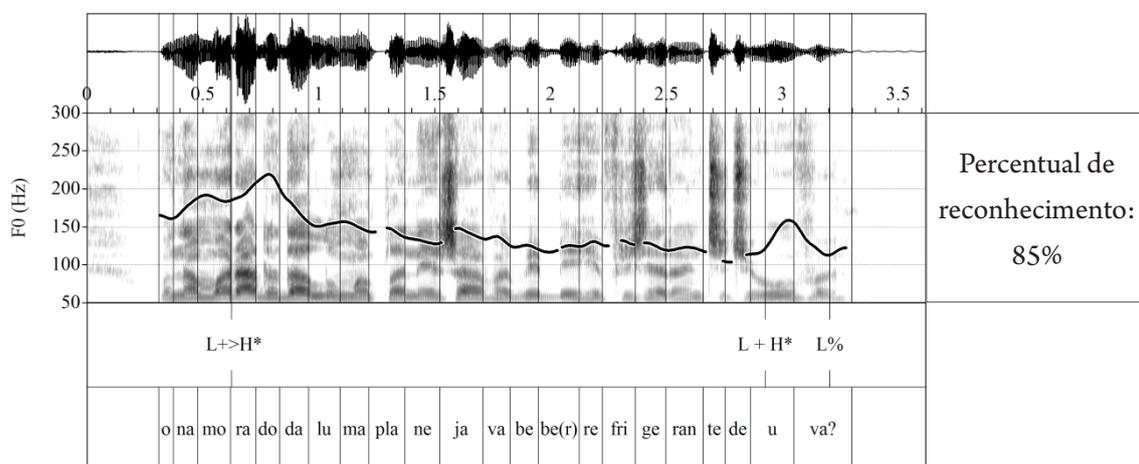
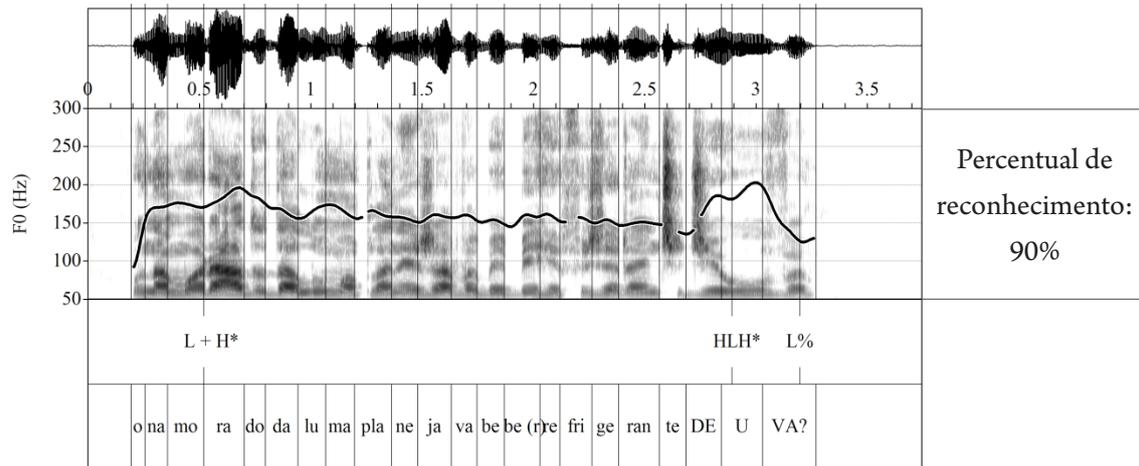


Figura 9. Forma de onda e curva de F0 suavizada (bandwidth 10 Hz) do enunciado *O namorado da Luma planejava beber refrigerante DE UVA?* pronunciado com foco sobre DE UVA pelo informante masculino; presença do pico geminado assinalada pelo retângulo em vermelho



Para o caso (ii), a não existência de pretônica, na focalização sobre QUENTE, fez com que fosse adotada uma outra estratégia: reforço duracional sobre a tônica da palavra em foco e alcance de maior amplitude do pico melódico nuclear, que passa de 172 Hz, no foco amplo (Figura 10), para 286 Hz, no foco estreito em posição nuclear (Figura 11):

Figura 10. Forma de onda e curva de F0 suavizada (bandwidth 10 Hz) do enunciado *Festa cheia costuma dar refresco quente?* pronunciado com foco amplo pelo informante masculino.

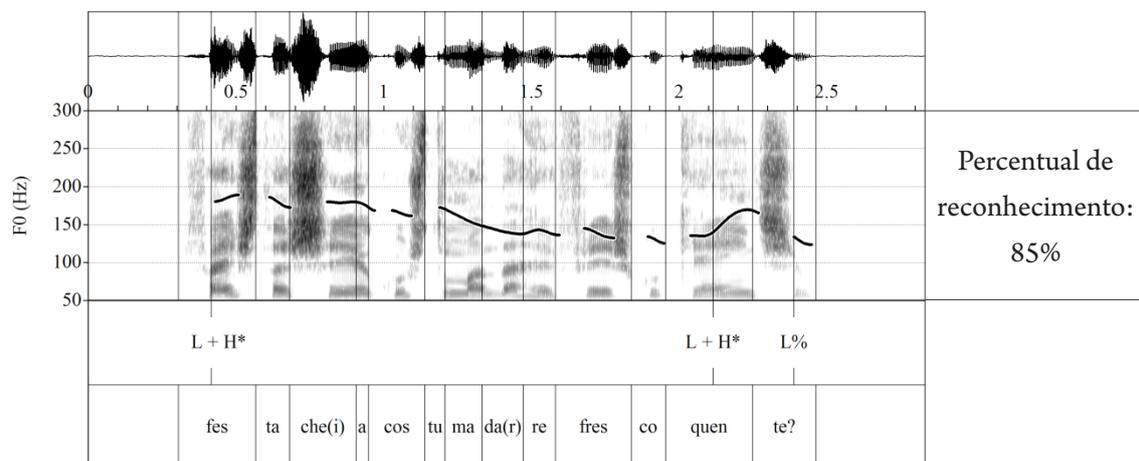
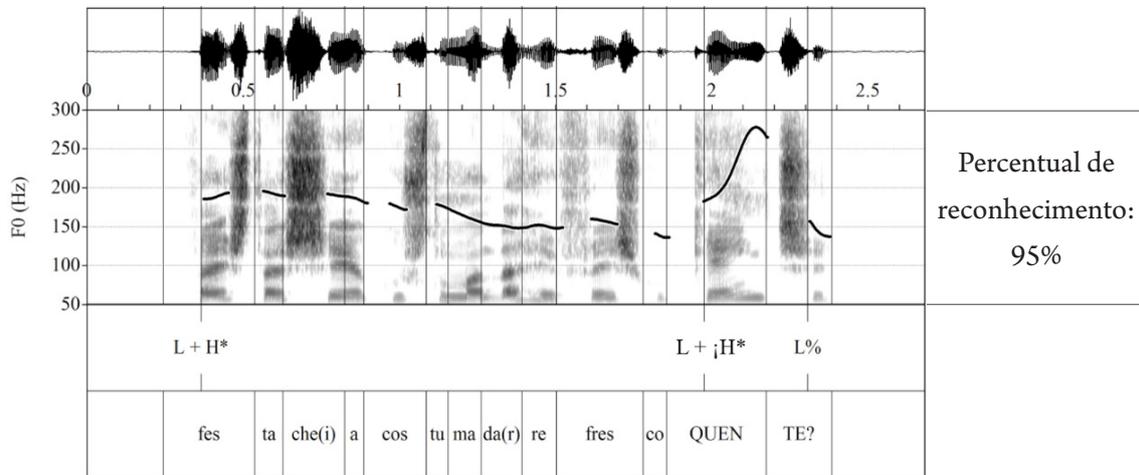


Figura 11. Forma de onda e curva de F0 suavizada (bandwidth 10 Hz) do enunciado *Festa cheia costuma dar refresco QUENTE?* pronunciado com foco sobre QUENTE pelo informante em masculino.



5.2.2 Focalização sobre cabeça de sintagma fonológico Φ

Nesta subseção, trataremos dos vocábulos focalizados que constituem cabeça de sintagma fonológico (Φ), mas que, diferentemente daqueles analisados anteriormente, não estão na fronteira de sintagma entonacional (I). A estrutura dos enunciados que compõem nosso *corpus* provê um par de vocábulos fonológicos para cada categoria criada a partir da quantidade de sílabas pretônicas: para vocábulo com duas ou mais sílabas pretônicas, tem-se DA MANUELA e DERRUBAR; para vocábulo com apenas uma sílaba pretônica DA LUMA e BEBER; para vocábulo sem pretônica, CHEIA e DAR.

Com relação às palavras fonológicas (ω) em posição de cabeça de Φ com duas ou mais pretônicas, caso de DA MANUELA e DERRUBAR, Figuras 12 e 13, respectivamente, podemos observar que a marcação do foco estreito ocorre por uma proeminência geminada sobre o vocábulo sob foco, proeminência essa que se acrescenta às duas proeminências *default* já mencionadas, a primeira em posição pré-nuclear, a segunda, de natureza modal, em posição nuclear. Soma-se a esta subida suplementar, caracterizada pelo pico geminado assinalado pelo retângulo vermelho nas Figuras 12 e 13, o aumento na duração das sílabas em que os picos melódicos do tom complexo HLH* estão localizados, ou seja, as sílabas tônica [é] e segunda pretônica a partir desta [ma], posição do acento secundário. O reforço duracional das sílabas, em relação à sua duração em contexto de foco amplo, aliado ao pico geminado são responsáveis pela marcação do foco estreito nestes dois vocábulos.

Figura 12. Forma de onda e curva de F0 suavizada (bandwidth 10 Hz) do enunciado *O primo DA MANUELA vai derrubar suco de tangerina?* pronunciado com foco em DA MANUELA pelo informante masculino; presença do pico geminado assinalada pelo retângulo em vermelho.

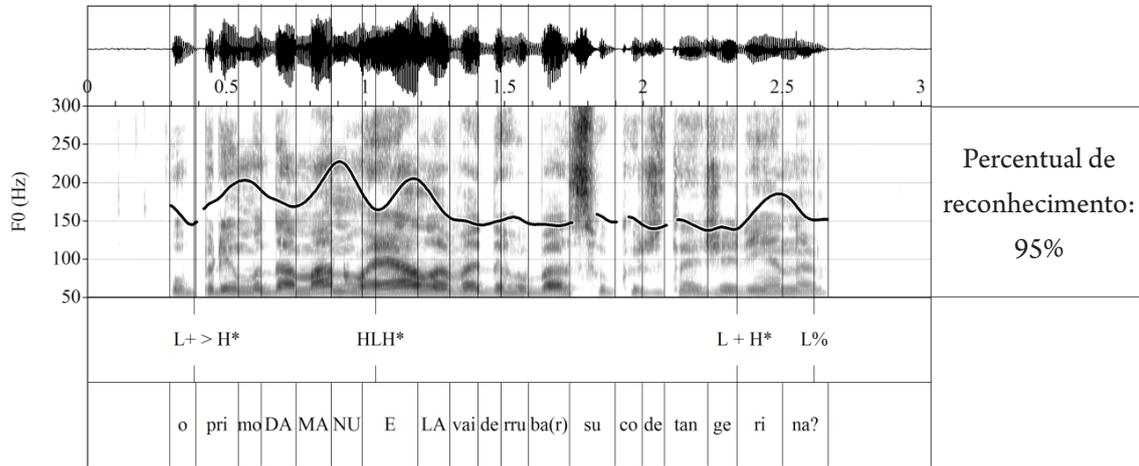
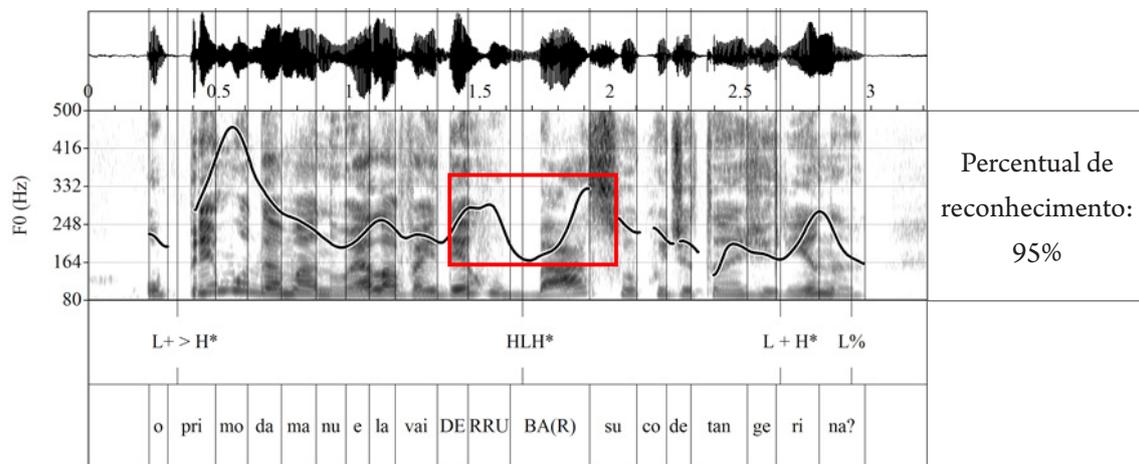


Figura 13. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O primo da Manuela vai DERRUBAR suco de tangerina?* pronunciado com foco em DERRUBAR pelo informante feminino; presença do pico geminado assinalada pelo retângulo em vermelho.



Para as cabeças de ϕ com apenas uma sílaba pretônica, DA LUMA e BEBER, ocorrem as mesmas estratégias na marcação do foco, porém, com algumas particularidades: quanto ao pico geminado da subida suplementar, pode ser percebida a presença do tom alto do acento secundário na única pretônica existente, configurando um pico geminado comprimido, mantendo-se o tom HLH* sobre a palavra fonológica, como descrito na focalização de vocábulo com uma pretônica em cabeça do sintagma entonacional (DE

UVA); quanto ao aumento da duração, esta estratégia ocorre apenas na sílaba tônica. Os padrões da focalização sobre DA LUMA e BEBER são ilustrados nas Figuras 14 e 15, respectivamente.

Figura 14. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O namorado DA LUMA planejava beber refrigerante de uva?* pronunciado com foco sobre DA LUMA pelo informante masculino; presença do pico geminado assinalada pelo retângulo em vermelho.

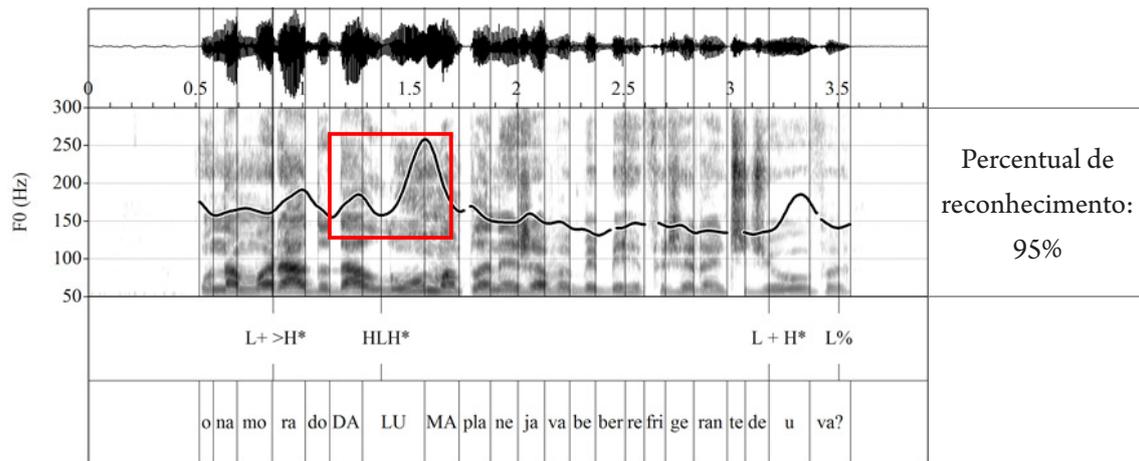
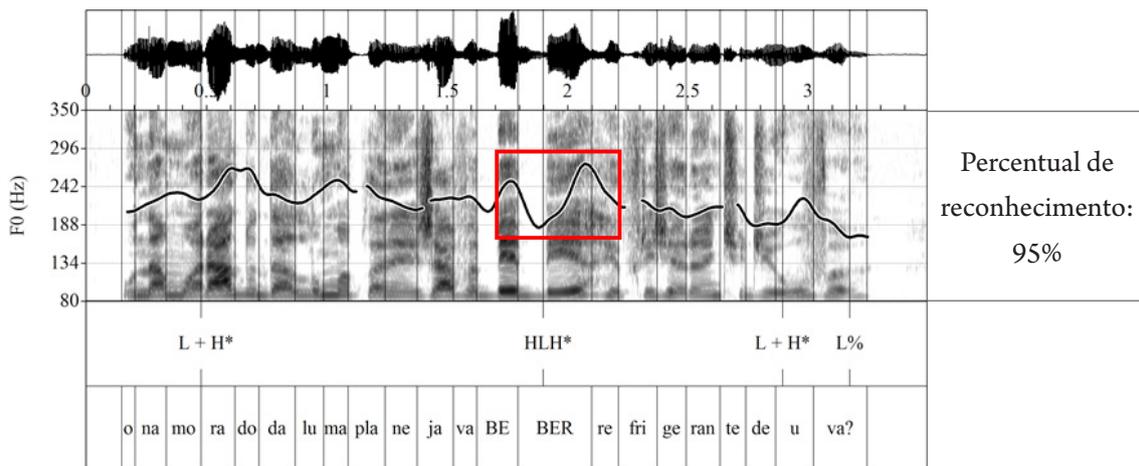


Figura 15. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O namorado da Luma planejava BEBER refrigerante de uva?* pronunciado com foco sobre BEBER pelo informante feminino; presença do pico geminado assinalada pelo retângulo em vermelho.



Assim, pode-se afirmar que o pico geminado parece ser um correlato robusto, já que ocorre até mesmo quando o acento secundário é inibido, em palavras com apenas

uma sílaba pretônica. No entanto, quando a cabeça de ϕ não apresenta pretônica, como nos casos de CHEIA e DAR (Figuras 16 e 17, respectivamente), a focalização mantém sua funcionalidade através de um pico suplementar simples e o reforço duracional sobre a tônica da palavra focalizada, apresentando um aumento de 83% da duração da sílaba tônica, quando comparada à duração da mesma sílaba em foco amplo, isto é, na ausência de foco. Em CHEIA, o informante feminino realizou, ainda, uma quebra temporal, a pausa após o vocábulo focalizado, na fronteira de ϕ , isolando o constituinte focalizado.

Figura 16. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *Festa CHEIA costuma dar refresco quente?* pronunciado com foco sobre CHEIA pelo informante feminino.

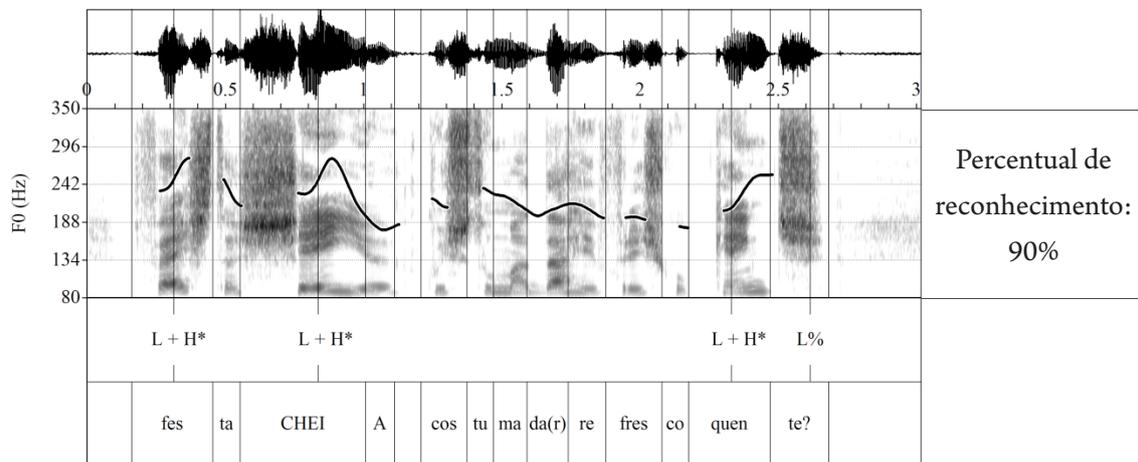
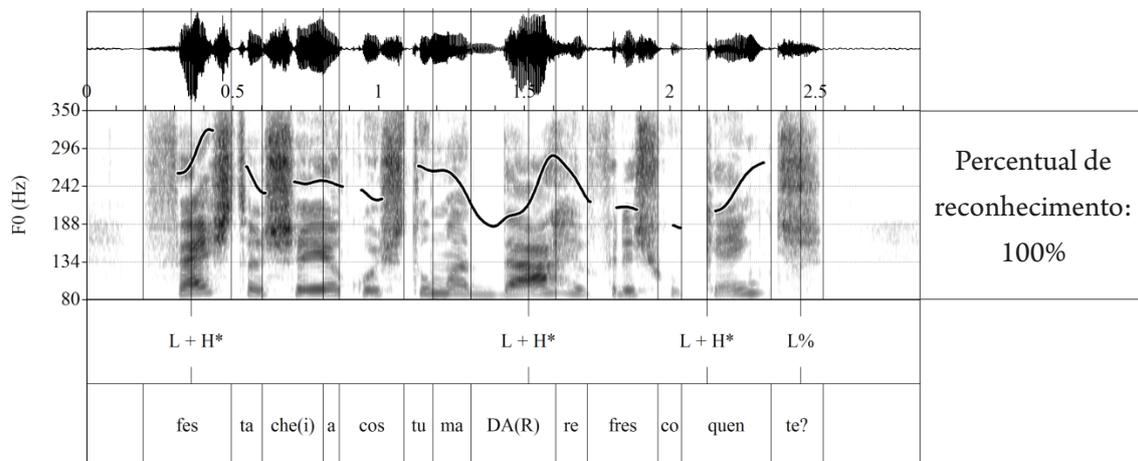


Figura 17. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *Festa cheia costuma DAR refresco quente?* pronunciado com foco sobre DAR pelo informante feminino.



5.2.3 Focalização sobre vocábulo à esquerda no sintagma fonológico, posição fraca de ϕ .

O último contexto de manifestação do foco estreito por nós analisado consiste na posição à esquerda do sintagma fonológico, posição fraca de ϕ . Para este caso, podemos distinguir três situações prosodicamente distintas, de acordo com a posição que o sintagma fonológico (ϕ) que contém a palavra em foco ocupa, dentro da estrutura hierárquica superior, isto é, do sintagma entonacional (I): (i) último ϕ de I, em que a posição forte de ϕ , à direita, apresenta a proeminência modal; (ii) primeiro ϕ de I, em que a posição fraca de ϕ , à esquerda, apresenta uma proeminência *default* por estar ao início de I; (iii) ϕ medial a I (à esquerda do ϕ interno ao I). Trataremos, portanto, deste contexto (vocábulo à esquerda de ϕ) discriminando essas três possibilidades.

(i) ϕ ao final de I (REFRIGERANTE, REFRESCO e SUCO)

Quando a palavra focalizada encontra-se em posição fraca, à esquerda, do último ϕ de I, observam-se duas estratégias principais, quais sejam, a presença do pico suplementar sobre o foco e o aumento da duração da sílaba tônica da palavra focalizada. Essas duas estratégias são recorrentes em todos os vocábulos, independentemente do número de sílabas pretônicas que apresentam. Em REFRIGERANTE, vocábulo com mais de duas sílabas pretônicas, percebe-se a manifestação do pico geminado no retângulo vermelho sobre a posição do acento secundário na produção do informante feminino (Figura 18). Nos casos de REFRESCO (Figura 18) e SUCO (Figura 19) ocorre um pico suplementar apenas.

Figura 18. Forma de onda e curva de F0 do enunciado *O namorado da Luma planejava beber REFRIGERANTE de uva?* pronunciado com foco sobre REFRIGERANTE pelo informante feminino; presença do pico geminado assinalada pelo retângulo em vermelho

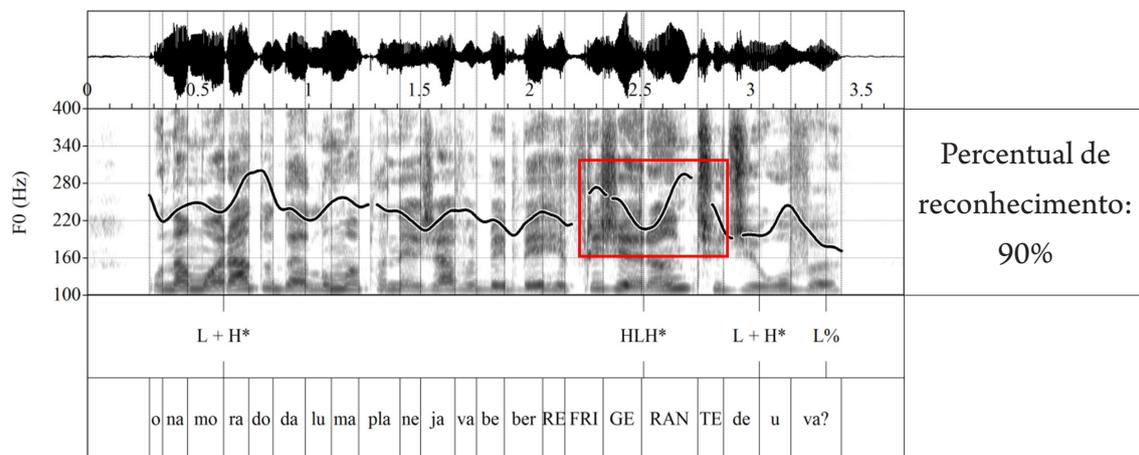


Figura 19. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *Festa cheia costuma dar REFRESCO quente?* pronunciado com foco sobre REFRESCO pelo informante feminino.

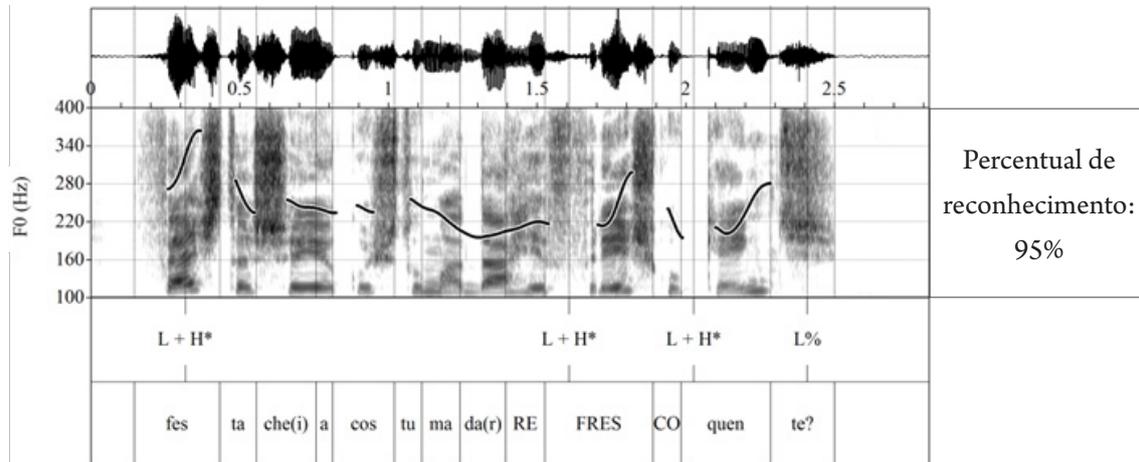
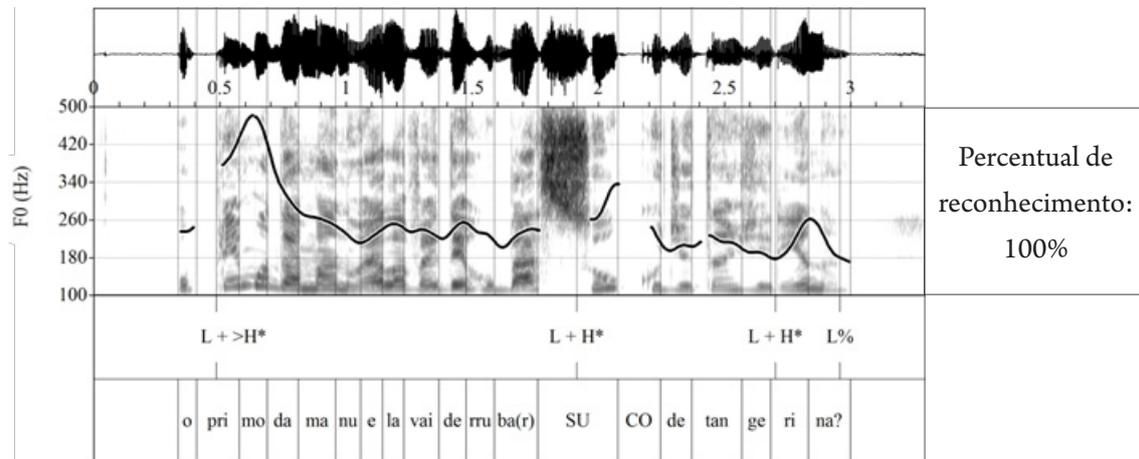


Figura 20. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O primo da Manuela vai derrubar SUCO de tangerina?* pronunciado com foco sobre SUCO pelo informante feminino.



(ii) ϕ ao início de I (O NAMORADO, O PRIMO e FESTA)

Ao início do sintagma entonacional, isto é, à esquerda de um ϕ ramificado, já se espera, como vimos, um movimento ascendente inicial, *default*, no pré-núcleo, por vezes bastante acentuado, mesmo na ausência de foco sobre esse constituinte (cf. por ex. Figuras 2, 4 e 13). Em vista disso, para a marcação do foco estreito sobre a palavra inicial em posição fraca, uma estratégia complexa entra em jogo. De um lado, temos a presença, além desse pico *default* sobre a primeira tônica, de um segundo pico, em geral de menor amplitude que o primeiro. Esse segundo pico, que não se confunde com o pico geminado (associado ao acento secundário), incide sobre o vocábulo no final do sintagma fonológico, que ocupa a cabeça

de ϕ (MORAES; CARNAVAL; COELHO, 2015), assinalando assim que há um foco sobre o primeiro ϕ do enunciado, e que ele recai sobre seu primeiro vocábulo (pico mais proeminente). Além disso, observa-se um alongamento sobre a tônica da palavra focalizada. Essa espécie de eco ou reduplicação do pico melódico (início e final do ϕ) ocorre tanto em palavras fonológicas que respeitam a condição mínima para a ocorrência do acento secundário, como O NAMORADO (Figura 21), quanto nas demais, como O PRIMO (Figura 22) e FESTA (Figura 23), como se pode observar na figuras a seguir. O mencionado reforço duracional sobre a tônica da palavra focalizada também se faz presente nesses três casos.

Figura 21. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O NAMORADO da Luma planejava beber refrigerante de uva?* pronunciado com foco sobre O NAMORADO pelo informante masculino; presença do pico geminado assinalada pelo retângulo em vermelho.

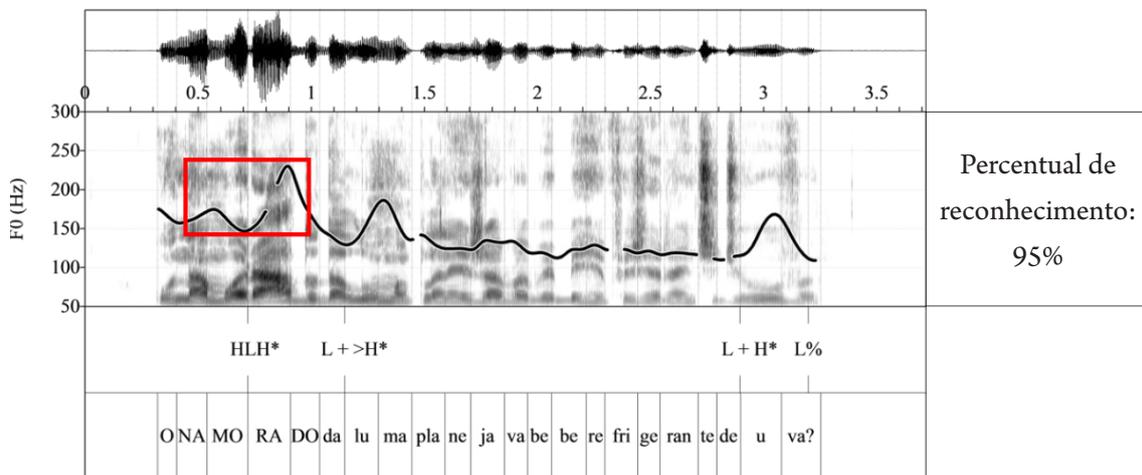


Figura 22. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O PRIMO da Manuela vai derrubar suco de tangerina?* pronunciado com foco sobre O PRIMO pelo informante masculino.

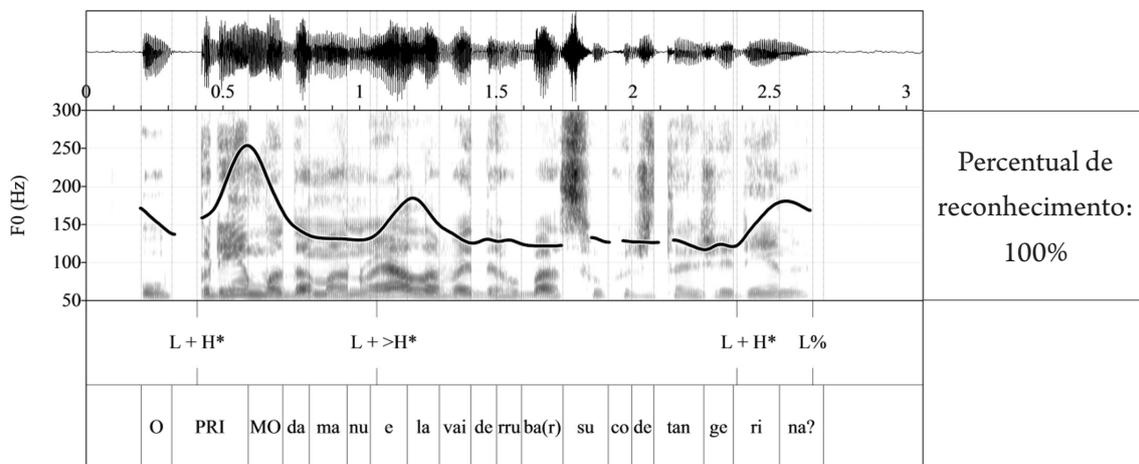
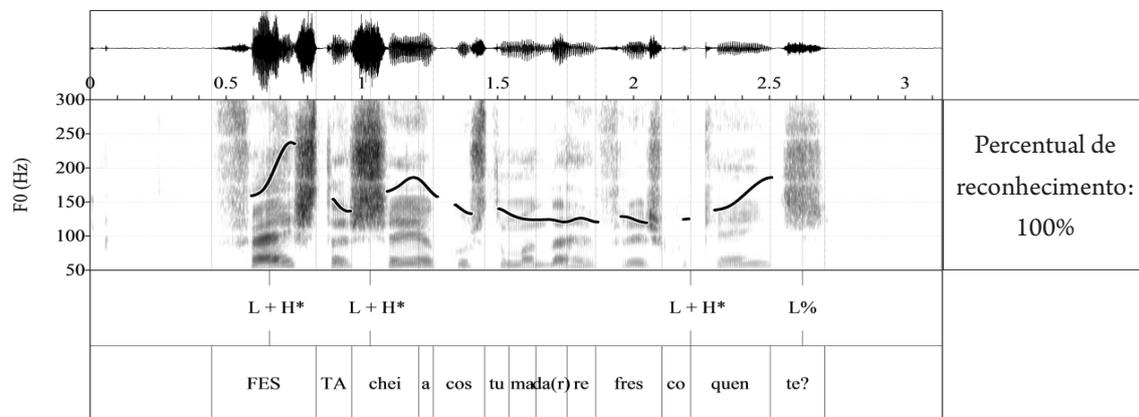


Figura 23. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *FESTA cheia costuma dar refresco quente?* pronunciado com foco sobre FESTA pelo informante masculino.



(iii) ϕ medial a I (PLANEJAVA, COSTUMA e VAI)

Quando o vocábulo focalizado ocupa a posição fraca do sintagma fonológico (ϕ) em posição medial do sintagma entonacional (I), ocorre sobre ele uma subida suplementar às subidas inicial *default* e final modal. Esta subida suplementar é caracterizada pelo pico geminado em palavra com duas pretônicas (PLANEJAVA – Figura 24), manifestando o acento secundário, e em palavra com apenas uma pretônica (COSTUMA – Figura 25), havendo a compressão do tom melódico complexo HLH* à estrutura textual do enunciado. Para a palavra sem pretônica (VAI – Figura 26), observa-se um pico simples sobre a palavra focalizada e, ainda, o alongamento significativo da sílaba tônica da palavra focalizada, o que ocorre nos três casos.

Figura 24. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *O namorado da Luma PLANEJAVA beber refrigerante de uva?* pronunciado com foco sobre PLANEJAVA pelo informante feminino; presença do pico geminado assinalada pelo retângulo em vermelho.

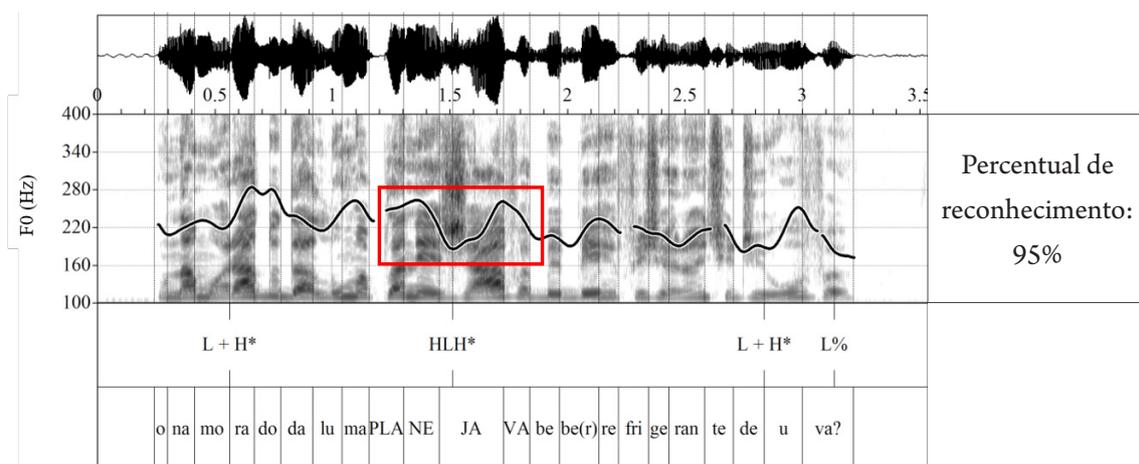


Figura 25. Forma de onda e curva de F0 suavizada (10 Hz) do enunciado *Festa cheia COSTUMA dar refresco quente?* pronunciado com foco sobre COSTUMA pelo informante feminino; presença do pico geminado assinalada pelo retângulo em vermelho.

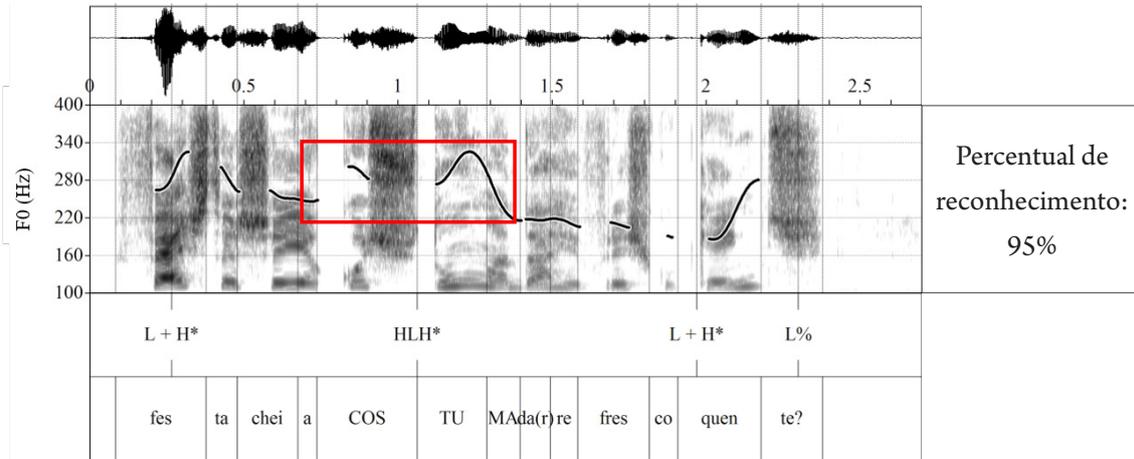
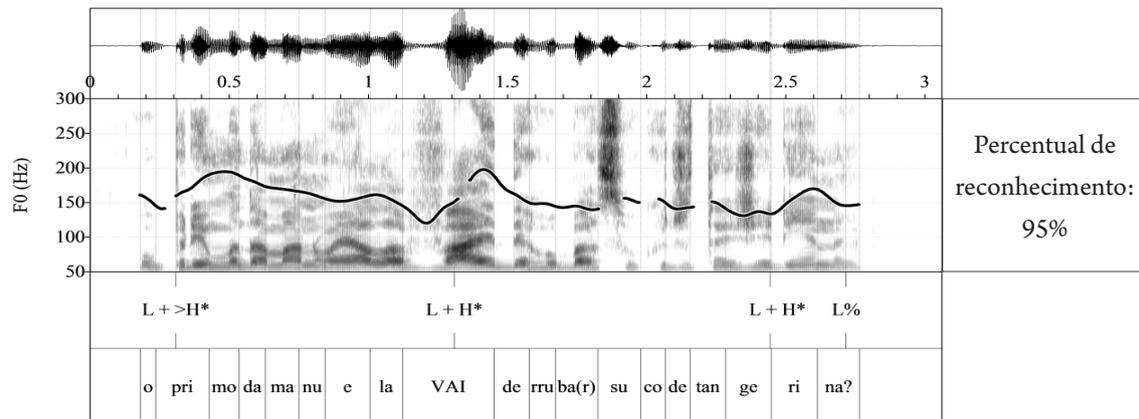


Figura 26. Forma de onda e curva de F0 do enunciado *O primo da Manuela VAI derrubar suco de tangerina?* pronunciado com foco sobre VAI pelo informante masculino.



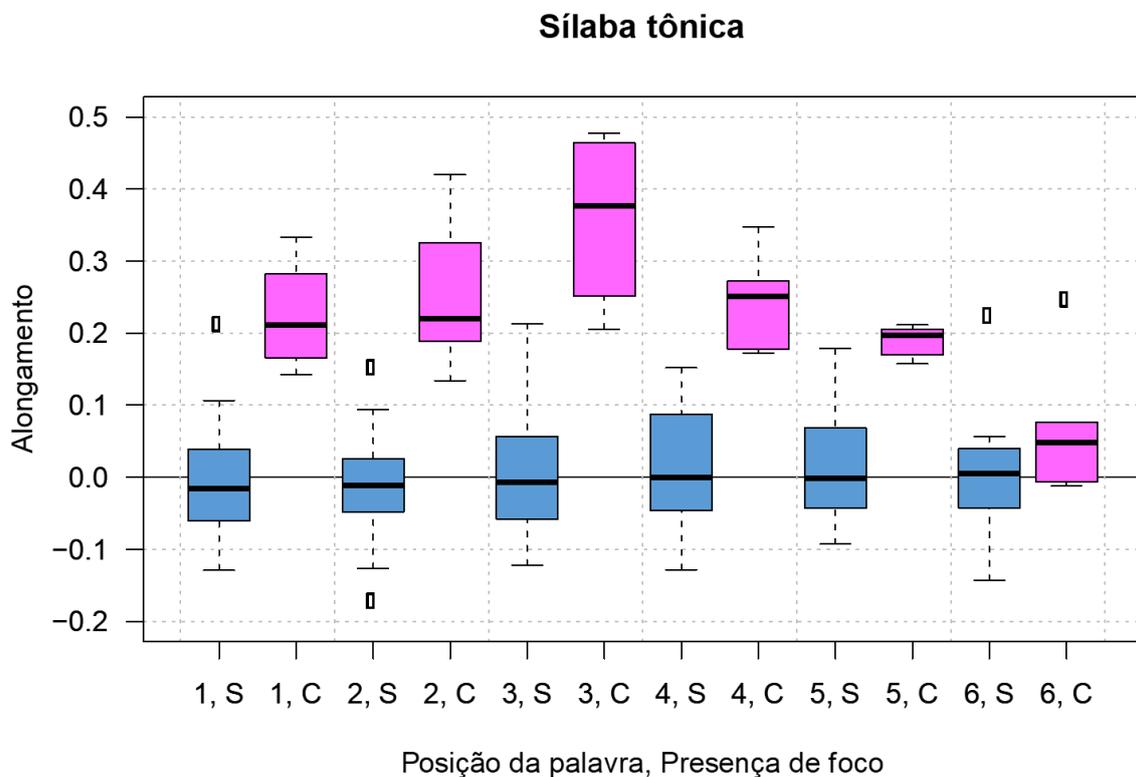
5.3 Análise da Duração

(i) Sílabas tônicas

A Figura 27 apresenta a distribuição das medidas de alongamento das sílabas tônicas em função da posição da palavra fonológica que contém a sílaba em questão e da presença ou não do foco estreito sobre essa palavra. Todas as sílabas tônicas dentro das palavras focalizadas apresentam aumento da duração em comparação com a sílaba correspondente no enunciado de foco amplo, com exceção da sílaba tônica da última palavra dos enunciados. As sílabas de palavras não focalizadas não apresentam aumento de duração em comparação com o enunciado de foco amplo. Um “Generalized Linear

Mixed Model” (GLMM — Modelo Misto Linear Generalizado, usando a função *lme()* da livreria *nlme* do programa R (PINHEIRO et al., 2018)) foi aplicado sobre as medidas de alongamento da duração das sílabas tônicas, utilizando como fatores fixos a presença ou não do foco sobre a palavra (2 níveis) e a posição da palavra dentro da frase (6 níveis), bem como a interação entre eles, e, ainda, como fator aleatório, o falante. O modelo mostrou que a presença do foco tem um efeito significativo ($F_{(1,203)} = 258, p < 0.001$) e explica a maior parte da variância no modelo. A interação entre os dois fatores (presença do foco e posição dentro da frase) também tem um efeito significativo ($F_{(5,203)} = 7.9, p < 0.001$), o que indica que a variação do aumento da duração nas sílabas tônicas ao longo da frase não é uniforme: o maior alongamento em contexto intermediário (posição 3), e menor (isto é, o não alongamento) em contexto final (posição 6) podem justificar esse resultado. De maneira global, a posição da palavra dentro da frase não tem um efeito significativo sobre o alongamento ($F_{(5,203)} = 2.1, p = 0.07$).

Figura 27. Distribuição das medidas de alongamento de sílabas tônicas com (C, em rosa) ou sem foco (S, em azul), em função da posição de sua palavra fonológica dentro da frase (1 a 6). Valores em 0 indicam que não houve aumento de duração da sílaba na frase com foco estreito em comparação à frase com foco amplo (ver texto), valores positivos indicam aumento e valores negativos compressão.

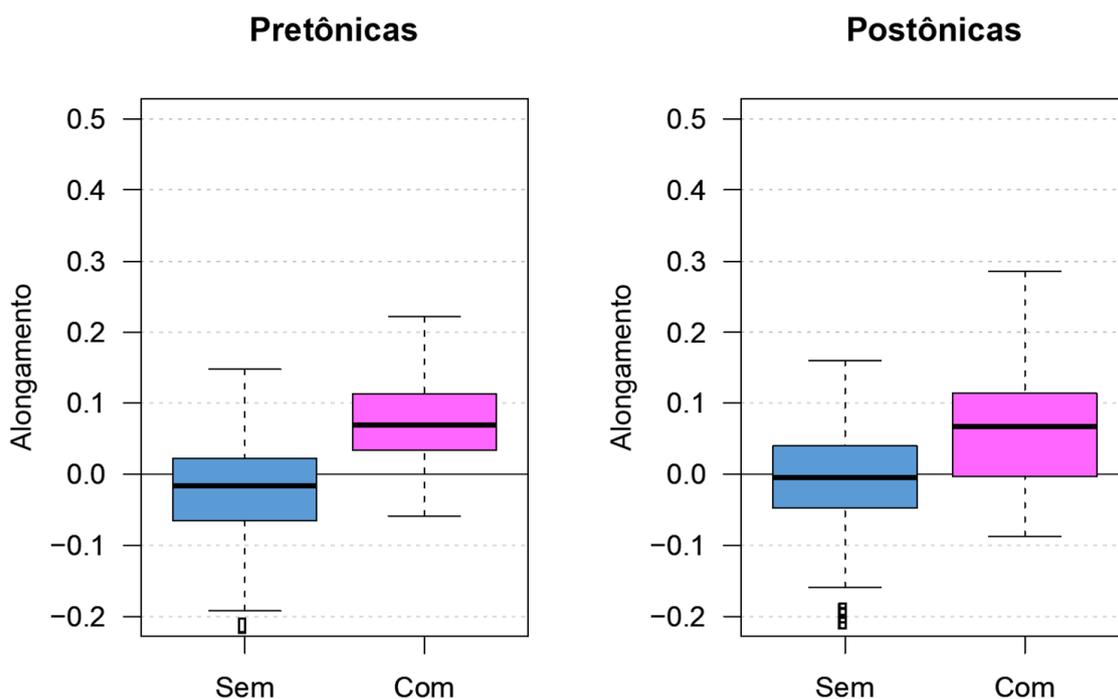


(ii) Sílabas pretônicas

Medidas similares foram realizadas para as sílabas pretônicas. Um GLMM foi aplicado na medida, composto dos mesmos dois fatores fixos utilizados para as sílabas tônicas (posição da palavra, presença ou não do foco) e mais um fator, descrevendo o tipo de sílaba pretônica (com três níveis: sílaba pretônica com um potencial acento secundário, sílaba pretônica única e as demais sílabas pretônicas que ocorressem). As interações entre esses três fatores foram incluídas dentro do modelo, com o fator Falante como aleatório. O modelo demonstrou que, também para as pretônicas, apenas a presença ou a ausência de foco sobre a sílaba apresenta um efeito significativo sobre o alongamento das sílabas ($F_{(1,227)} = 71, p < 0.001$). Os demais fatores e as interações entre eles não atuam significativamente sobre este alongamento, isto é, o aumento de duração não depende significativamente do tipo de sílaba pretônica, apresentando esta um potencial acento secundário ou não.

A Figura 28, à esquerda, apresenta as distribuições das medidas de alongamento por sílabas pretônicas com ou sem foco. O alongamento médio das sílabas pretônicas em palavras com foco é aproximadamente de 20% - a ser comparado ao alongamento observado sobre as sílabas tônicas de palavras focalizadas, que se situa em torno de 60% (média de 0.2) em relação às mesmas sílabas com foco amplo.

Figura 28. Distribuição das medidas de alongamento das sílabas pretônicas (à esquerda) e postônicas (à direita), com (em rosa) ou sem foco (em azul). Valores em 0 indicam que não houve aumento de duração da sílaba na frase com foco estreito em comparação à frase com foco amplo (ver texto), valores positivos indicam aumento e valores negativos compressão.



(iii) Sílabas postônicas

Outro modelo GLMM foi aplicado para as medidas de alongamento observadas nas sílabas postônicas, com dois fatores fixos (presença ou não de foco, posição da palavra - nesse *corpus*, não há sílabas postônicas de palavras em posição 4, não sendo, portanto, consideradas no modelo) e o fator aleatório, o Falante. Para as postônicas, o modelo também demonstrou que a presença de foco é um fator significativo no alongamento ($F_{(1, 129)} = 22, p < 0.001$) e este explica a maior parte da variância (ver Figura 28, à direita). A interação significativa ($F_{(4, 129)} = 3.6, p < 0.01$) entre a posição da palavra e a presença de foco indica que o alongamento da postônica varia também com a posição dentro da frase, mesmo que seu efeito explique uma parte reduzida da variância. Essa interação é relacionada à ausência de alongamento das sílabas postônicas com foco em posição final dos enunciados

6 Considerações finais

Com a aplicação dos testes perceptivos, foi possível conferir o bom reconhecimento dos padrões melódicos da focalização estreita em interrogativas totais no PB, observando os seguintes contextos: vocábulo final do enunciado, coincidindo com cabeça de I; vocábulo cabeça de sintagma fonológico (ϕ), posição mais forte; vocábulo que não constitui cabeça de ϕ , situando-se em posição fraca, à esquerda, podendo o sintagma fonológico estar ao início, meio ou final do enunciado. Ao considerar a atuação do processo nos domínios da palavra fonológica (ω), baseando-se na quantidade de sílabas pretônicas da palavra focalizada, o presente trabalho procurou averiguar a relevância da manifestação do acento secundário na marcação do foco estreito e quais as estratégias adotadas em sua ausência.

A partir dessa sistematização, foi verificado que o acento secundário no foco interrogativo apresenta como marca formal o pico geminado de tom complexo HLH* e que este atua na marcação do foco estreito em todos os contextos analisados, quando a condição mínima para manifestação do acento secundário é respeitada (ao menos duas sílabas pretônicas). Quando esta condição não é cumprida, parece ocorrer, de forma recorrente, a compressão do pico geminado em duas sílabas contíguas, em palavras que apresentam apenas uma pretônica, ajustando-se a melodia ao material textual. Portanto, a manifestação melódica do acento secundário na forma de um pico geminado constitui um importante correlato da marcação do foco estreito em interrogativas totais no PB. Esse correlato atuaria em conjunto com o alongamento significativo da sílaba tônica em contextos de

foco estreito em comparação à duração das mesmas sílabas em contexto de foco amplo, configurando uma estratégia complexa de focalização. Quando o vocábulo em foco não apresenta pretônica alguma e, portanto, não pode manifestar o acento secundário, sequer de forma comprimida, esse aumento da duração parece atuar de modo expressivo para a marcação do foco estreito. No entanto, essa afirmação não se confirma aqui de forma conclusiva, já que, em nosso *corpus*, há poucas ocorrências de palavras sem sílaba pretônica, o que não torna viável avaliar estatisticamente a significância da diferença entre o alongamento das sílabas tônicas em palavras focalizadas com e sem pretônica⁶.

Em relação ao parâmetro duração, foram considerados dois fatores fixos: a presença ou não de foco e a posição da palavra dentro da frase. Para o primeiro fator, a análise estatística demonstrou, de forma geral, que é significativo o alongamento das sílabas tônicas em palavras focalizadas em comparação a essas mesmas sílabas em contexto sem foco (foco amplo). Foi também constatada uma variação na duração entre sílabas pretônicas com um potencial acento secundário e as demais sílabas pretônicas. No entanto, a significância dessa variação não se comprovou estatisticamente em nosso *corpus*, devido ao número reduzido de palavras com ocorrência de acento secundário. Portanto, para resultados mais conclusivos em relação à variação de duração da sílaba com acento secundário, é necessária a análise de um *corpus* mais amplo, construído de modo a viabilizar a análise do efeito potencial do acento secundário sobre a duração em contexto de foco. Em relação à posição da palavra focalizada dentro da frase, foi observada uma variação da duração das sílabas entre as posições do enunciado, que também não pode ser apontada como significativa, devido ao reduzido número de ocorrências em análise.

Vale, por fim, ressaltar que, mesmo em posições prosodicamente fracas, a F0 marca com bastante eficiência o elemento em foco. Isso é especialmente relevante no interior do primeiro sintagma fonológico, onde já há, na posição pré-nuclear, uma proeminência melódica *default*, e, sobretudo, quando não há espaço para a manifestação do acento secundário. Ainda assim, a distinção entre o foco sobre o primeiro elemento do sintagma *vs.* o segundo (por ex. “FESTA cheia” *vs.* “festa CHEIA”) e, ainda, sua oposição em relação ao foco amplo é marcada e percebida com sucesso. Nesses três casos, a primeira tônica terá um pico (*default*); se o foco for amplo, ele será o único pico ao longo do sintagma fonológico, e a F0 deverá ir decaindo ao longo do enunciado, até a tônica final. Se, por outro lado, o foco estreito incidir sobre o primeiro vocábulo do sintagma fonológico, haverá ao

⁶ Excluídos os casos de *clash* acentual, foco em “suco”, e de palavra ao início do enunciado, “festa”, percebe-se que as sílabas tônicas de palavras focalizadas sem pretônica apresentam alongamento de 83%, enquanto as sílabas tônicas de palavras focalizadas com pretônica têm o alongamento de 66%.

menos dois picos, um sobre cada uma das tônicas (por ex. em FES_{ta} e em CHE_{ia}), sendo o primeiro mais acentuado (atingindo um nível mais alto).⁷

Referências

ARANTES, P. Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala. Tese de Doutorado, Universidade Estadual de Campinas, Campinas, 2010.

_____.; BARBOSA, P. A. Acentuação secundária em Português Brasileiro à luz do modelo dinâmico do ritmo: um estudo piloto. *Anais do I Congresso Internacional de Fonética e Fonologia / VII Congresso Nacional de Fonética e Fonologia*. Belo Horizonte, 2002.

_____.; _____. F1 and Spectral Correlates of Secondary Stress in Brazilian Portuguese. In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (eds.). *Proceedings of the Speech Prosody 2008 Conference*. Campinas, Brazil: Editora RG/CNPq, 2008, p. 559-562.

AVESANI, C.; VAYRA, M. Broad, narrow and contrastive focus. In: Florentine Italian. *Proceedings of the 15 th ICPhS*, Barcelona, 2003.

BARBOSA, P. A. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, v. 49, n. 9, p. 725-742, 2007.

BELIÃO, J. How prosody and syntax are mapping: a study of synchronization and congruence. PhD Thesis, Université Paris Ouest Nanterre la Défense, 2016.

BISOL, L. O acento e o pé métrico binário. *Cadernos de Estudos Lingüísticos*, v. 22, p. 69-80, 1992.

_____. O acento e o pé binário. *Letras de Hoje* 29 (4), p. 25-36, 1994.

CARNAVAL, M. Foco informacional e foco contrastivo no português do Brasil: uma abordagem prosódica. Dissertação de Mestrado em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.

COLLISCHONN, G. Um estudo do acento secundário em português. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.

_____. Acento secundário em português. *Letras de Hoje* 29 (4) p. 43-53, 1994.

DI CRISTO, A. *De la microprosodie à l'Intonosyntaxe*. Aix-en-Provence: Univerté de Provence, 1985.

FERNANDES-SVARTMAN, F. Acento secundário, atribuição tonal e ênfase em português brasileiro (PB). *Estudos Linguísticos*, 38 (1), p. 47-58, 2009.

⁷ Se couber um acento secundário por haver mais de uma pretônica, como em namorado, teremos três picos NAmoRAdo da LUma; o mesmo se aplica quando a segunda palavra é que recebe o foco, como em o PRImo da MAnuEla. Por fim, se o foco estreito estiver sobre o segundo vocábulo do sintagma, haverá igualmente dois picos sobre as duas tônicas, sendo agora o segundo deles mais proeminente. Além do comportamento melódico diferenciado, observa-se, nesses casos, o alongamento da tônica do vocábulo sob o foco.

_____.; ABAURRE, M. B. M.; GONZÁLEZ-LÓPEZ, V. A.; BIANCHI, M.C.C. Secondary stress, intensity and fundamental frequency in Brazilian Portuguese. *Journal of Portuguese Linguistics*, v. 11, n. 2, 51–67, 2012.

FROTA, S. *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing, 2000.

_____.; CASTELO, J.; CRUZ, M.; BARROS, N.; CRESPO-SENDRA, V.; SILVESTRE, A. P. Melodia ou texto? Estratégias de acomodação entre melodia e texto no Português. In.: *Revista Diadorim*, v. 17, p. 12-33, 2015.

HAAN, J. *Speaking of Questions. An exploration of Dutch question intonation*. Utrecht: LOT, 2002.

JUN, S. (Ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 2005.

_____. (Ed.) *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 2014.

KRUIJFF-KORBAYOVÁ, I.; STEEDMAN, M. Discourse and information structure. *Journal of Logic, Language and Information*, v. 12, n. 3, p. 249-259, 2003.

LADD, D. R. *Intonational Phonology*. Cambridge: CUP, second edition, 2008.

LEE, S. Acento secundário do PB. *Letras de Hoje*, v. 37, n. 1, p. 149-162, 2002.

MITTMAN, M. O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro. Tese de Doutorado em Linguística, Faculdade de Letras, UFMG, 2012.

NESPOR, M.; VOGEL, I. *Prosodic Phonology*. Dordrecht: Foris, 1986.

MORAES, J., 2003a. A manifestação fonética do pé métrico, *Letras de Hoje* 134, dez. 2003, p. 147-162.

_____, 2003b. Secondary stress in Brazilian Portuguese: perceptual and acoustical evidence. In: Proceedings of the XV International Congress of Phonetic Science [Barcelona 3-9 August 2003], Glenelg North, Australia: Casual, 2003, p. 2063-2066.

_____. Variações em torno de tema e rema. *Cadernos do CNLF [IX Congresso Nacional de Linguística e Filologia]* v. IX, n. 17, 2006, p. 279-289.

_____., CARNAVAL, M.; COELHO, A. A manifestação prosódica do foco em interrogativas totais no Português do Brasil e sua percepção. *Revista Virtual de Estudos da Linguagem – ReVEL*, v. 13, n. 10, p. 170-194, nov. 2015.

PINHEIRO J.; BATES D.; DEBROY S.; SARKAR, D. and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-137, <https://CRAN.R-project.org/package=nlme>. 2018.

ROSSI, A. Qual é a natureza do acento secundário no português brasileiro? *Cadernos - Centro Universitário São Camilo*, v. 4, n. 1, p. 77-92, 1998.

SANDALO, M.; F.; ABAURRE, M. B. M.; MANDEL, A.; GALVES, C. M. C. Secondary stress in two varieties of Portuguese and the Sotaq optimality based computer program. *Probus*, v. 18. n. 1, p. 97-125, 2006.

SITYAEV, D.; HOUSE, J. Phonetic and phonological correlates of broad, narrow and contrastive focus in English. Proceedings of the 15th ICPPhS, Barcelona, 2003.

VALLDUVÍ, E. *The Informational Component*. New York : Garland, 1992.

_____. *Information packaging: a survey*. Report prepared for WOPIS (Word Order, Prosody, and Information Structure). University of Edinburgh, 2003.



Data de submissão: 15/03/2018

Data de aceite: 21/11/2018

“PRA CONQUISTAR A GAROTA...”: DESGARRAMENTO E PROSÓDIA NAS VARIEDADES BRASILEIRA E LUSITANA DO PORTUGUÊS

“TO WIN THE GIRL...”: DETACHMENT AND PROSODY
IN BRAZILIAN AND EUROPEAN PORTUGUESE

Aline Ponciano dos Santos Silvestre | [Lattes](#) | aponcianossilvestre@letras.ufrj.br
Universidade Federal do Rio de Janeiro

Resumo: Neste artigo, estuda-se o comportamento prosódico de orações adverbiais *desgarradas* no Português Brasileiro (PB) e no Português Europeu (PE), sendo utilizados o arcabouço teórico da Fonologia Prosódica (NESPOR; VOGEL, 1994) e da Fonologia Entoacional (PIERREHUMBERT 1980, LADD 1996). Foram analisados 1800 dados de leitura (900 de cada variedade do português) e feitas aferições de duas pistas prosódicas: contorno melódico e duração das sílabas no fim do sintagma entoacional (IP). Os resultados revelam que o *desgarramento* na língua falada parece ser licenciado, primordialmente, tanto em PB quanto em PE, pela maior duração nas sílabas finais do IP. Para o PB, além da variação fonética dada pelo comportamento duracional das últimas sílabas do IP, o *desgarramento* é caracterizado por um padrão melódico diferente do verificado nas orações adverbiais anexadas à oração núcleo (majoritariamente, L+H*L% para as *não desgarradas* e L+H*H% para as *desgarradas*), o que sugere o fato de, para além da variação fonética, o fenômeno constituir um padrão fonológico diverso no português brasileiro.

Palavras-chave: Prosódia; *Desgarramento*; Variedades entoacionais.

Abstract: In this article, we study the prosodic behavior of *detached* adverbial clauses in Brazilian Portuguese (PB) and European Portuguese (PE). We use the theoretical framework of prosodic phonology (NESPOR; VOGEL, 1994) and Intonational Phonology (PIERREHUMBERT 1980, LADD 1996). 1800 read data (900 of each Portuguese variety) were analyzed and measurements were made on two prosodic parameters: melodic contour and duration at the end of the Intonational Phrase (IP). The results reveal that the *detachment* in the oral language seems to be primarily licensed, in PB as in PE, because the speakers insert longer duration in the final syllables of the IP. For PB, in addition to the phonetic variation given by the durational behavior of the last syllables of the IP, the *detachment* is characterized by a different melodic pattern from the adverbial clauses attached to the main sentence (L+H*L% for the sentences with the main clause and L+H*H% for the detached ones), which suggests that the phenomenon constitutes a different phonological pattern in Brazilian Portuguese.

Keywords: Prosody; *Detachment*; Intonational varieties.

Introdução

Este artigo reúne alguns achados da tese de Silvestre (2017) e objetiva caracterizar, prosodicamente, orações adverbiais *desgarradas* em duas variedades do português. Mais especificamente, numa comparação entre estruturas que ocorrem juntamente à oração núcleo e estruturas adverbiais que ocorrem sozinhas, almeja-se descrever pistas prosódicas que permitem a compreensão de uma oração que, sendo chamada *subordinada* na tradição gramatical, ao contrário do que tal tradição postula, existe na língua sem ser dependente da denominada oração “principal”.

A existência das chamadas orações *desgarradas* foi defendida por Decat (1999, 2011), com base em uma análise funcional discursiva. Atendo-se primordialmente à análise de dados escritos, a autora percebe o fenômeno do *desgarramento* como uma estratégia a serviço da produção textual, através da qual o autor do texto produziria sequências como “Esse caso com a modelo Lilian Ramos realmente foi uma tragédia. “Apesar de Itamar ser um senhor solteiro e o ambiente ter sido de Carnaval” (DECAT, 2011, p. 33), em que há o uso de pontuação não canônica como estratégia de focalização. Ressalte-se aqui, entretanto, a presença da chamada oração “principal” na sequência.

Além do *desgarramento* majoritariamente analisado por Decat (1999, 2011) em seus estudos, a autora menciona, em sua descrição do fenômeno, um tipo de oração *desgarrada* que denota a clara possibilidade de orações adverbiais ocorrerem totalmente soltas. Exemplos disso são as orações “Se eu ganhasse na Sena!” ou a que dá título a esse trabalho – “Pra conquistar a garota...”, adverbiais sozinhas, soltas, *desgarradas*, mas completamente interpretáveis dentro de determinado contexto comunicativo.

Decat (1999, 2011), portanto, trata como *desgarradas* duas estruturas que aqui – porque procederemos a uma descrição baseada em princípios da fonologia – são essencialmente diferentes. E é ao segundo tipo de estrutura *desgarrada* mencionada anteriormente – estruturas semelhantes à do título – a que nos ateremos, distinguindo-as das anteriores e batizando-as de *desgarradas totais*. Analogamente à tradição dos estudos prosódicos, que nomeiam como *questões totais* perguntas para as quais a resposta pode ser apenas “sim” ou “não”, uma vez que contêm toda a informação desejada, chamaremos de *desgarradas totais* as orações adverbiais em que a oração núcleo não é recuperável textualmente, pelo fato de, tal qual as referidas *questões totais*, serem adverbiais que possuem, sozinhas, toda a informação necessária à sua interpretação. Sendo o *desgarramento* nosso tema, referir-nos-emos, em nossa análise, às orações formalmente anexadas à cláusula núcleo como *não desgarradas*.

1 Enquadramento teórico

Divergindo do que é usualmente postulado pela tradição gramatical, Decat (1999, 2011), fundamentada numa análise funcional-discursiva, defende a necessidade de verificação do tipo de dependência (forma, sentido, pragmática) considerado para a definição do *status* dependente ou não das cláusulas. A autora afirma que

[...] na caracterização da dependência de uma cláusula a outra, o parâmetro formal apresenta-se como o mais utilizado. Entretanto, conforme ressalta Thompson (1984), uma análise que fique presa exclusivamente a indicadores formais terá, forçosamente, de considerar a cláusula subordinada como dependente. (DECAT, 2011, p. 24)

Decat (2011) aponta, então, a distinção entre dois grupos de subordinadas: 1) encaixadas: aquelas que são cláusulas dependentes, estruturalmente integradas, e que desempenham um papel gramatical em *constituência* com um item lexical, grupo no qual se encontram as tradicionalmente chamadas *substantivas* e *adjetivas restritivas*; e 2) hipotáticas: aquelas que são cláusulas dependentes e que representam opções organizacionais para os falantes, das quais emergem *proposições relacionais* (inferências), podendo constituir, elas mesmas, *unidades de informação* à parte, grupo no qual se encontram as tradicionais *adjetivas explicativas* e as *adverbiais*.

Ainda segundo considerações de Decat (2011), as estruturas de hipotaxe, cláusulas menos dependentes e que, portanto, podem formar uma *unidade de informação* por si, estariam propensas ao *desgarramento*, ou seja, teriam a possibilidade de ocorrerem, sintaticamente, independentes na língua:

[...] a noção de “unidade de informação” está correlacionada com a ocorrência isolada de cláusulas subordinadas. Caracterizando-se como opções do discurso, servindo a objetivos comunicativo-interacionais, tais cláusulas “desgarram-se” porque constituem **unidades de informação à parte**, o que as reveste de um menor grau de dependência, tanto formal quanto semântica, chegando mesmo a se identificarem como cláusulas tidas como independentes, à maneira de alguns tipos de coordenadas. A dependência que se estabelece, nesses casos, será pragmático-discursiva. (DECAT, 2011, p. 42, grifos nossos)

Não há dúvidas quanto à consistência do *status* pragmático-discursivo que dá vida às orações *desgarradas*, como bem preveem e descrevem as análises funcionalistas. Do ponto de vista da análise funcional-discursiva das *desgarradas* feita por Decat (1999, 2011), parte-se, portanto, da ideia sintático-pragmática de que as adverbiais são cláusulas hipo-

táticas, menos subordinadas formalmente, mas dependentes do contexto, para, depois, iniciar-se a discussão sobre a *unidade de informação* (cf. CHAFE, 1980) e, indiretamente, perceber-se a relevância de cláusulas *desgarradas* como um constituinte fonológico, através da observação de comportamentos prosódicos como pausa e entoação, comportamentos esses pouco explorados nas análises funcionalistas. Na análise aqui empreendida, entretanto, um outro ponto de vista se coloca, o qual tem como partida uma assunção fonológica: a oração *desgarrada total* é um sintagma entoacional (IP) e um enunciado (U). Uma vez que IP e U são, respectivamente, domínios de um contorno melódico e de uma unidade de sentido, a oração *desgarrada total* traz consigo, necessariamente, uma caracterização prosódica própria, que necessita ser descrita, e um sentido em si. A relação com a sintaxe passa, então, a ser secundária, limitada à necessidade de construção dos constituintes prosódicos.

Estabelecido nosso ponto de vista, as subseções a seguir dedicam-se às definições teóricas de ordem fonológica que guiarão nossos procedimentos de análise.

1.1 A Fonologia Prosódica

A teoria proposta por Nespor e Vogel (1986,1994) questiona a adequação da teoria gerativa inicial que limitou a interação da fonologia com o restante da gramática a uma inter-relação com a sintaxe. As autoras argumentam que o componente fonológico da gramática não deve ser visto de forma homogênea e sim como “um subconjunto de subsistemas em interconexão, cada um governado com princípios próprios” (NESPOR; VOGEL 1994, p. 13).

Desse modo, segundo os postulados da teoria prosódica, a corrente fônica está dividida em fragmentos hierarquicamente organizados - os constituintes prosódicos - os quais estão delimitados por diferentes indícios, que abrangem desde modificações segmentais em si até mudanças fonéticas mais sutis. Os referidos constituintes prosódicos, distribuídos de forma decrescente na hierarquia, são o enunciado fonológico (U – *Utterance*), o sintagma entoacional (IP – *Intonational Phrase*), o sintagma fonológico (PhP – *Phonological Phrase*), o grupo clítico (CG – *Clitical Group*), a palavra fonológica (PW – *Prosodic Word*), o pé (F - *Foot*) e a sílaba (Syl – *Syllable*).

No decorrer dos estudos alicerçados em postulados da Fonologia Prosódica, a existência de processos fonológicos que podem ser aplicados ou impedidos devido à relação com os limites dos domínios prosódicos, sejam processos segmentais (como o sândi e a elisão) ou suprasegmentais (como a retração do acento e a entoação), tem sido utiliza-

da como prova da distribuição hierárquica dos constituintes (HAYES; LAHIRI 1991, TRUCKENBRODT 1995, FROTA 2000, FROTA; VIGÁRIO 2000, TENANI 2002).

Para além dos processos estritamente fonológicos que licenciam a referida distribuição hierárquica, Nespor e Vogel (1994) afirmam que os constituintes da hierarquia prosódica proporcionam estruturas relevantes para o primeiro nível de processamento da percepção da fala, o *parsing* inicial, fornecendo ao ouvinte a base para a reconstrução da estrutura sintática e para a compreensão da mensagem transmitida por uma dada sequência (NESPOR; VOGEL 1994, p. 287). Baseando-se nas sugestões de Selkirk (1978) e nas afirmações de Nespor e Vogel (1983a, 1983b), que se utilizaram de dados perceptivos em seus estudos, as autoras salientam que

[...] não são os constituintes sintáticos, mas os constituintes prosódicos os que proporcionam a informação relevante na primeira etapa de processamento de uma sequência de fala. Isso não quer dizer que a estrutura sintática seja irrelevante, mas que só é relevante indiretamente, uma vez que só se faz referência à informação sintática na construção dos constituintes prosódicos que se situam acima da palavra prosódica. Da afirmação de que são os constituintes prosódicos, e não os sintáticos, os que proporcionam as unidades relevantes para o nível inicial de processamento se segue que toda distinção sintática não refletida na estrutura prosódica não pode ser captada nesse nível de percepção. (NESPOR; VOGEL, 1994, p. 288, tradução nossa)

Ao desenvolver uma proposta prosódica para explicar casos de desambiguação – e tal fato nos interessa particularmente porque as adverbiais *desgarradas totais* que estudamos têm interpretação diversa, mas possuem exatamente a mesma estrutura sintática das adverbiais anexadas à oração núcleo – Nespor e Vogel (1994) declaram que os casos de maior possibilidade de desambiguação são aqueles em que há estruturas prosódicas diferentes no nível do IP, asseverando que

as orações que se podem desambiguar são aquelas em que os diferentes significados correspondem a diferentes estruturas prosódicas. Ao contrário, as orações em que os diferentes significados têm a mesma estrutura prosódica não são desambigáveis, independentemente de sua estrutura sintática. (NESPOR; VOGEL, 1994, p. 293, tradução nossa)

Tendo por base tais informações, neste trabalho, assim como em Tenani (2002), nossa análise será pautada na observação dos três níveis mais altos da hierarquia prosódica - U, IP e PhP - uma vez que são esses os níveis largamente descritos como responsáveis

pela percepção e diferenciação de estruturas. Por esta razão, acreditamos serem também esses níveis os mais importantes para que se possam verificar as marcas prosódicas caracterizadoras do *desgarramento* nas variedades estudadas.

Importa mencionar que a adoção das abordagens postuladas pela Fonologia Prosódica justifica-se bastante por almejarmos uma comparação da estrutura entoacional associada aos domínios prosódicos em estruturas *desgarradas* e não *desgarradas* no PB e no PE. Isso porque, como não há outros estudos prosódicos sobre o *desgarramento*, é a abordagem teórica utilizada que nos permite uma comparação coerente entre as variedades, a fim de que, com base em trabalhos que tratam da estrutura prosódica do português (para o PE: FROTA; VIGÁRIO 2000, FROTA; VIGÁRIO 2001, VIGÁRIO 2003, FERNANDES 2007, SEVERINO 2011, BARROS 2014. PARA O PB: FROTA; VIGÁRIO 2000, TENANI 2002, FERNANDES 2007, SERRA 2009, FONSECA 2010), seja observado se as pistas prosódicas caracterizadoras do fenômeno em estudo são variações fonéticas ou se constituem um padrão fonológico diferente.

1.2 A Fonologia Entoacional

Além de considerarmos a hierarquia prosódica e seus constituintes, lançamos mão das abordagens feitas pelo modelo autossegmental e métrico (AM) da Fonologia Entoacional, postuladas por Pierrehumbert (1980), Ladd (2008), entre outros.

O modelo AM prevê uma organização fonológica própria para a entoação, interpretando-a como uma sequência de eventos tonais localizados, diretamente relacionados com a acentuação e com fronteiras de domínios prosódicos. Portanto, pode-se presumir que a estrutura prosódica, explicitada na seção anterior, condiciona, de algum modo, a estrutura entoacional. Dentro do modelo AM, assume-se que a constituição das melodias se dá por sequências de dois tipos de tons, apenas (altos [H] e baixos [L]), e são também dois os tipos de eventos tonais suficientes para descrevê-las: os acentos tonais (*pitch accents*) e os tons de fronteira (*boundary tones*).

Além de acentos tonais e de tons de fronteira, suficientes para a descrição fonológica da maioria das línguas, há a possibilidade de existir um acento intermediário (*intermediate phrase*), chamado acento frasal. Para o português, o trabalho de Tenani e Fernandes-Svartman (2008) dá indícios da possibilidade de ser alocado um tom de fronteira ao PhP em sentenças focalizadas, o que indica a existência de um acento frasal no PB.

A conjugação do modelo hierárquico e do modelo AM é feita, para o Português, em trabalhos como os de Frota (2000, 2002, 2003), Frota e Vigário (2000), Tenani (2002),

Viana e Frota (2007), Fernandes (2007), Serra (2009), Fonseca (2010), Severino (2011), Cruz e Frota (2011), Silvestre (2012), Barros (2014), Frota et al. (2015), Castelo (2016), entre muitos outros, que recobrem diversas variedades do PB e do PE.

Descritas brevemente as bases teóricas que regem nosso estudo, passaremos à descrição de nosso *corpus* e da metodologia de análise que permitiu a descrição do *desgarramento* em dados de fala do PB e do PE.

2 Corpus e Metodologia

O *corpus* foi montado a fim de que pudéssemos proceder a uma análise comparativa de orações adverbiais *não desgarradas* e *desgarradas totais*, em busca da descrição prosódica do fenômeno que nos motiva – o *desgarramento*. Deste modo, as orações que serviram de base para o estudo foram obtidas através de gravações de um *corpus* de leitura, no qual foram descritas situações em que o uso de orações adverbiais *desgarradas* ou *não desgarradas* é possível, o que permitiu a comparação de trechos lexicalmente idênticos.

Todas as situações foram apresentadas em *slides* e foi solicitado às informantes que, após pensados os contextos, somente as orações-alvo fossem lidas. A pesquisa contou com dez informantes do sexo feminino, cinco oriundas da região do Grande Rio – alunas de pós-graduação em Letras na Universidade Federal do Rio de Janeiro, e cinco oriundas da região de Lisboa – alunas de pós-graduação em Letras na Universidade de Lisboa, com idades entre 23 e 36 anos.

O *corpus*, como um todo, foi composto de 30 orações adverbiais base: 15 que fazem parte de estruturas complexas, com orações adverbiais anexadas à oração núcleo - *não desgarradas*, e outras 15, correspondentes, *desgarradas totais*. Cada oração foi lida três vezes por todas as informantes, a fim de que pudéssemos confirmar a regularidade das características prosódicas observadas.

Sendo o sintagma entoacional nossa unidade básica de análise, há, em nossos dados, orações adverbiais *desgarradas totais* e *não desgarradas* de estruturas diferentes: orações menores, de nove sílabas, sem ramificação no último PhP; e orações maiores, com treze sílabas, em que o último PhP é ramificado. Tais estruturas foram pensadas a fim de que se pudessem testar as hipóteses concernentes à influência do tamanho do IP ou do peso fonológico na inserção das pistas prosódicas que caracterizam o *desgarramento*, hipóteses essas que, de acordo com nossos objetivos, não serão aqui detalhadas, mas podem ser encontradas em Silvestre (2017).

As 900 orações adverbiais para cada variedade do português aqui estudada assim

distribuídas: 225 *não desgarradas* sem ramificação no último PhP (15 frases x 5 informantes x 3 repetições), 225 *não desgarradas* com último PhP ramificado (15 frases x 5 informantes x 3 repetições), 225 *desgarradas totais* sem ramificação no último PhP (15 frases x 5 informantes x 3 repetições) e 225 *desgarradas totais* com último PhP ramificado (15 frases x 5 informantes x 3 repetições).

A seguir, exemplificamos como se deu o processo de obtenção das orações, sendo solicitado às informantes a imaginação dos contextos (indicados por [C:]), com posterior leitura das sentenças em destaque. O mesmo contexto foi apresentado quatro vezes, de forma randomizada, para que fosse feita a leitura das orações *desgarradas* e *não desgarradas*, com ramificação ou não no último PhP:

[C: Você conversa com seu irmão sobre tudo o que o seu amigo, o Rodrigo, fazia para agradar à primeira namorada. Seu irmão começa a descrever os ursinhos, as flores, os chocolates que Rodrigo comprava... E você comenta:]

Pra conquistar a garota, gastava mundos e fundos. (*Não desgarrada*, sem ramificação no último PhP)

Pra conquistar a garota... (*Desgarrada*, sem ramificação no último PhP)

Pra conquistar a garota desejada, gastava mundos e fundos. (*Não desgarrada*, com ramificação no último PhP)

Pra conquistar a garota desejada... (*Desgarrada*, com ramificação no último PhP)

De acordo com a mesma configuração dos exemplos anteriores, todas as orações destacadas para a leitura, aqui apresentadas em sua ramificação ideal, foram as seguintes:

- Estruturas com nove sílabas – sem ramificação no último PhP:

- (1) [[Se a Joelma]PhP [a ganhou]PhP]IP
- (2) [[Se o Ricardo]PhP [desejasse]PhP]IP
- (3) [[Se o Diogo]PhP [conseguisse]PhP]IP
- (4) [[Quando o Fábio]PhP [me chamasse]PhP]IP
- (5) [[Quando a Ana]PhP [apontasse]PhP]IP
- (6) [[Quando a Carla]PhP [imagina]PhP]IP
- (7) [[Já que o Lázaro] PhP [desejava] PhP]IP
- (8) [[Já que o Leandro] PhP [o procura] PhP]IP

- (9) [[Já que a Marina] PhP [gostaria] PhP]IP
- (10) [[Pra aprovar] PhP [os alunos] PhP]IP
- (11) [[Pra conquistar] PhP [a garota] PhP]IP
- (12) [[Pra enviar] PhP [os pedidos] PhP]IP
- (13) [[Embora a Vera] PhP [suplicasse] PhP]IP
- (14) [[Embora a Lúcia] PhP [o tentasse] PhP]IP
- (15) [[Embora a Carmen] PhP [a quisesse] PhP]IP

Estruturas com treze sílabas – com ramificação no último PhP:

- (16) [[Se a Joelma]PhP [ganhasse na loteria]PhP]IP
- (17) [[Se o Ricardo]PhP [desejasse o emprego]PhP]IP
- (18) [[Se o Diogo]PhP [conseguisse o trabalho]PhP]IP
- (19) [[Quando o Fábio]PhP [chamasse ao escritório]PhP]IP
- (20) [[Quando a Ana] PhP [apontasse a janela] PhP]IP
- (21) [[Quando a Carla] PhP [imagina as tragédias] PhP]IP
- (22) [[Já que Lázaro] PhP [desejava o perigo] PhP]IP
- (23) [[Já que Leandro] PhP [procura o empregado] PhP]IP
- (24) [[Já que Marina] PhP [gostaria dos enfeites] PhP]IP
- (25) [[Pra aprovar] PhP [os alunos esforçados] PhP]IP
- (26) [[Pra conquistar] PhP [a garota desejada] PhP]IP
- (27) [[Pra enviar] PhP [os pedidos requeridos] PhP]IP
- (28) [[Embora Vera] PhP [suplicasse aos juízes] PhP]IP
- (29) [[Embora Lúcia] PhP [tentasse o resultado] PhP]IP
- (30) [[Embora Carmen] PhP [quisesse a recompensa] PhP]IP

2.1 Processos de análise do *corpus*

Para a efetiva notação prosódica, utilizamos sistema P_TOBI, postulado, para o português, por Vianna e Frota (2007) e por Frota (2014), sistema esse baseado no TOBI inglês (TO para *Tones* e BI para *Break Indices*), o qual foi desenvolvido dentro da teoria AM (BECKMAN; HISCHBERG; SHATTUCK-HUFNAGEL 2005) e propõe o alinha-

mento do contorno de F0 a uma série de camadas. O sistema de notação P_TOBI é feito com o auxílio do programa PRAAT e inclui três camadas: uma para associação tonal, uma para a transcrição ortográfica e uma para a anotação de fronteiras prosódicas.

A fim de descrever o comportamento prosódico de orações *desgarradas* no PB e no PE, aqui foram considerados um parâmetro acústico de natureza duracional – alongamento silábico – e um parâmetro acústico de natureza melódica – a modulação da frequência fundamental.

Procedemos à descrição do contorno melódico presente no primeiro PhP de todos os IPs e a aferição da duração concentrou-se na palavra nuclear (pré-fronteira direita) dos IPs constituídos por orações adverbiais, *desgarradas* e *não desgarradas*. Isso porque, como demonstram trabalhos construídos sob a mesma base teórica (TENANI 2002, FERNANDES 2007, SERRA 2009, FONSECA 2010 – para o PB; e FROTA 2000, FROTA; VIGÁRIO 2001, SEVERINO 2011, BARROS 2014 – para o PE), é ela – a fronteira final – o principal *locus* para a inserção de características prosódicas capazes de diferenciar estruturas em português.

No que concerne especificamente ao comportamento duracional das sílabas, alinhamo-nos ao que descreve Serra (2009) em sua análise sobre o fraseamento do PB, e esperamos que o alongamento se revele de forma mais expressiva nas sílabas tônica e pós-tônica finais do que na sílaba pré-tônica, já que esta se encontra mais distante da fronteira. Seguindo os passos de análise da autora, assumimos que

[...] para se observar se houve ou não alongamento, deve-se contrastar a sílaba “suspeita” de alongamento, de preferência com outra “não suspeita”, pois se as duas alongam juntas, não se consegue evidenciar, em termos relativos, se houve ou não o referido alongamento. A rigor, qualquer sílaba que não alongue poderia ser escolhida para servir de referência, isto é, para contrastar com a postônica. A pretônica 1 é uma boa candidata, pois, (i) além de não alongar (razão principal), (ii) é conhecida a relação “ideal” de sua duração com as demais (tônica e postônica) fora do contexto fronteira (cf Moraes 1995), e (iii) é ela, como a postônica, uma sílaba átona, ou seja, têm elas durações em princípio mais próximas entre si (do que entre a postônica e a tônica, por exemplo), o que faz com mais frequência ser positivo o índice do alongamento, evidenciando-o melhor. Assim, é mais elegante dizer que a postônica alonga de X% em relação à pretônica, do que dizer que a postônica “desalonga” menos do que o esperado em relação à tônica, por exemplo. (SERRA 2009, p. 74)

Uma vez que analisamos estruturas lexicalmente idênticas para a comparação de orações não *desgarradas* e *desgarradas*, a observação da existência de alongamento se deu

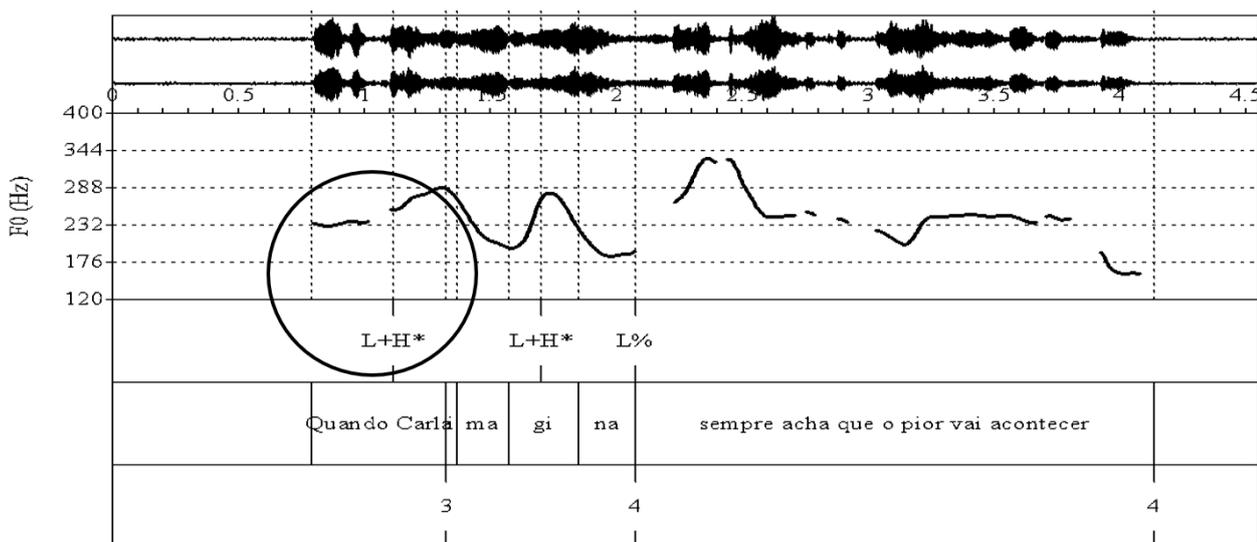
em duas etapas: 1) através da comparação interssilábica, em que medimos a duração das três sílabas finais da palavra nuclear e descrevemos, separadamente para cada tipo oracional, o percentual de aumento ou descenso das sílabas pré-tônica e pós-tônica em relação à tônica; 2) através da comparação interoracional, em que realizamos a comparação da duração média das sílabas da palavra nuclear em *orações desgarradas* e *não desgarradas*.

A anotação das modulações de F0 (com base no P_ToBI) foi feita no programa PRAAT (BOERSMA; WEENICK, 2015) e todos os contornos observados – nas fronteiras inicial e final do IP – foram descritos em planilhas do Excel para que, posteriormente, fosse feita a contagem percentual dos contornos predominantemente encontrados.

3 Resultados

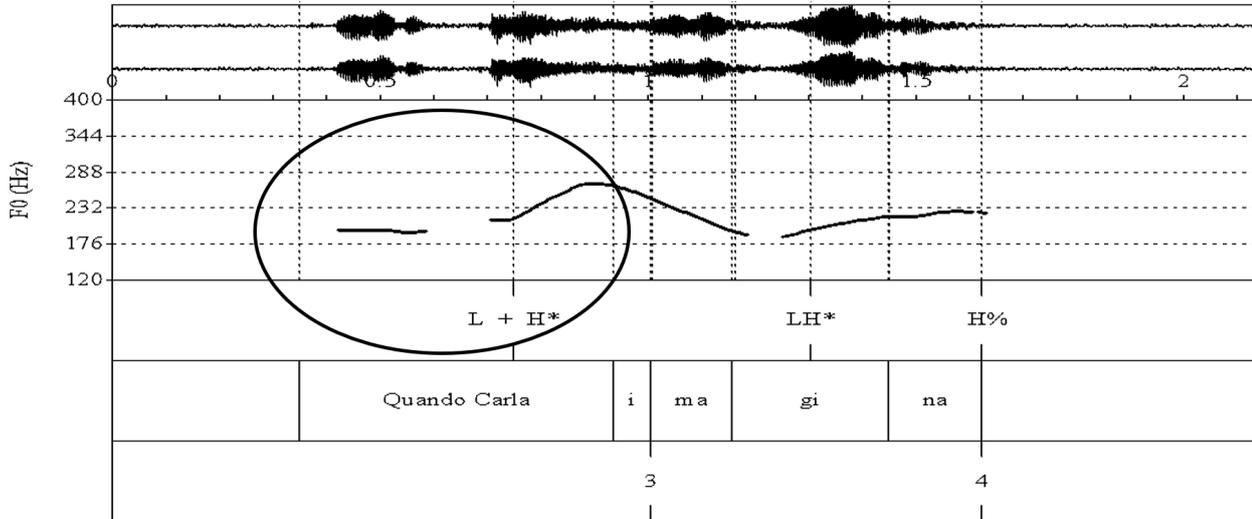
A análise do contorno melódico inicial de *orações desgarradas* e *não desgarradas* revelou que os mesmos contornos foram observados no elemento proeminente do primeiro PhP, seja em *orações não desgarradas* ou em *orações desgarradas totais*, não havendo, assim, diferenças concernentes à associação tonal inicial que caracterizem o fenômeno do *desgarramento*. Entretanto, as variedades analisadas diferem pelo fato de o acento bitonal L+H* ser mais produtivo nos dados brasileiros ao passo que a utilização do acento H+L* é mais efetiva nos dados portugueses, como mostram as figuras seguintes (Figura 2 a Figura 5) e revelam os Gráficos 1 e 2:

Figura 2. Contorno L+H* predominante no PhP inicial de orações não desgarradas do PB



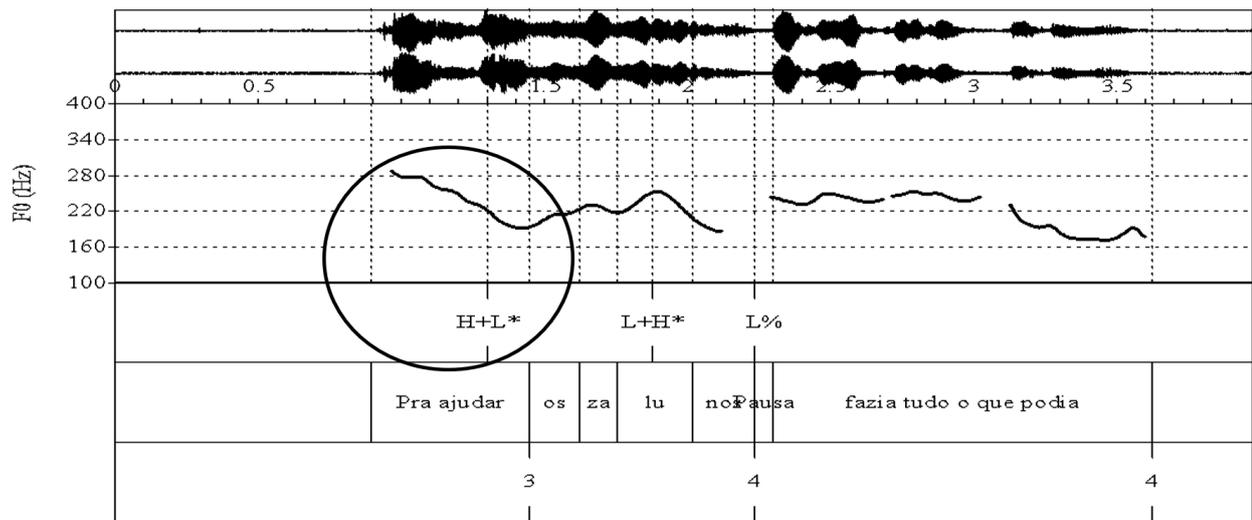
Fonte: Silvestre (2017).

Figura 3. Contorno L+H* predominante no PhP inicial de orações desgarradas do PB



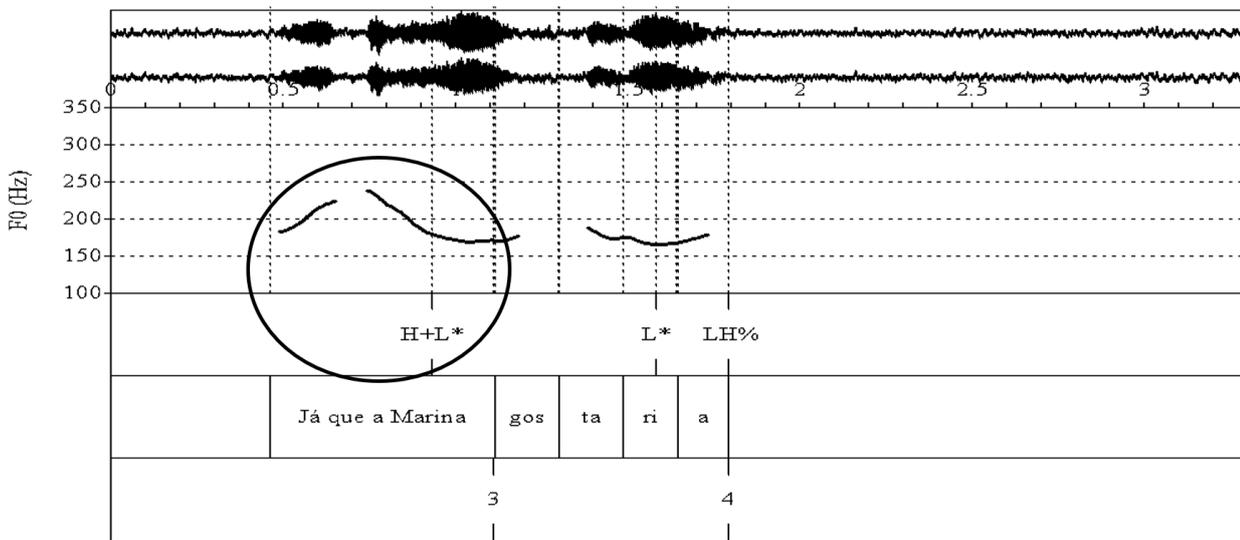
Fonte: Silvestre (2017).

Figura 4. Contorno H+L* predominante no PhP inicial de orações não desgarradas do PE



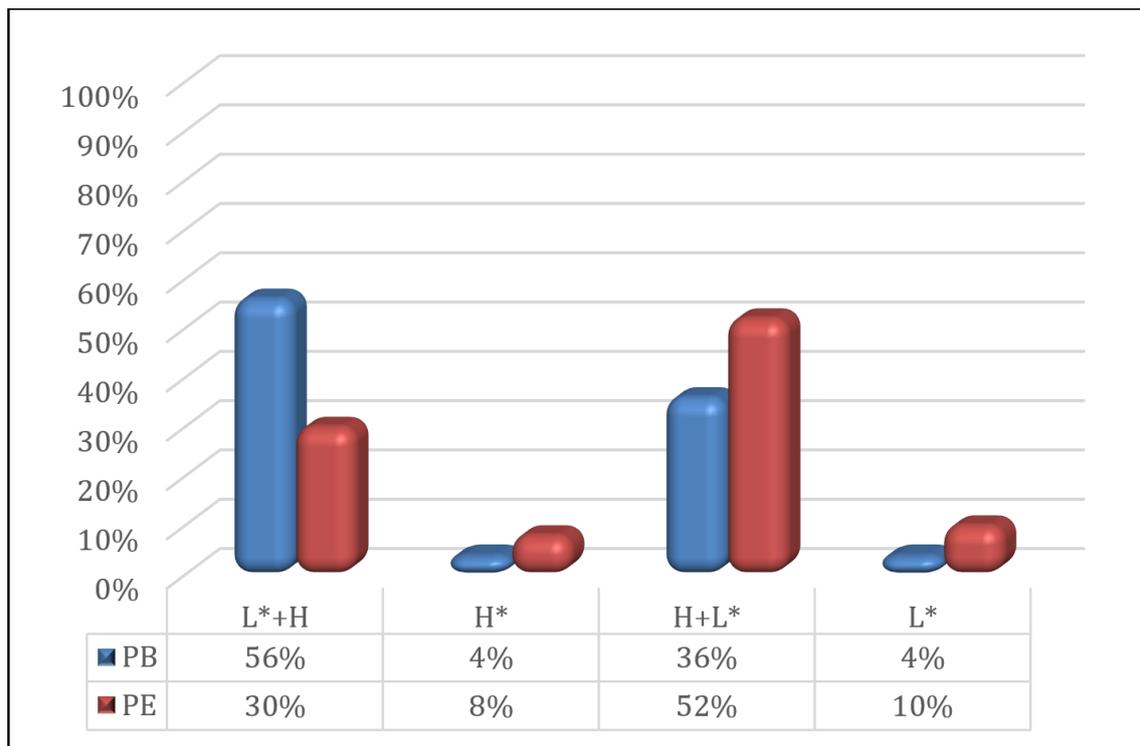
Fonte: Silvestre (2017).

Figura 5. Contorno H+L* predominante no PhP inicial de orações desgarradas do PE



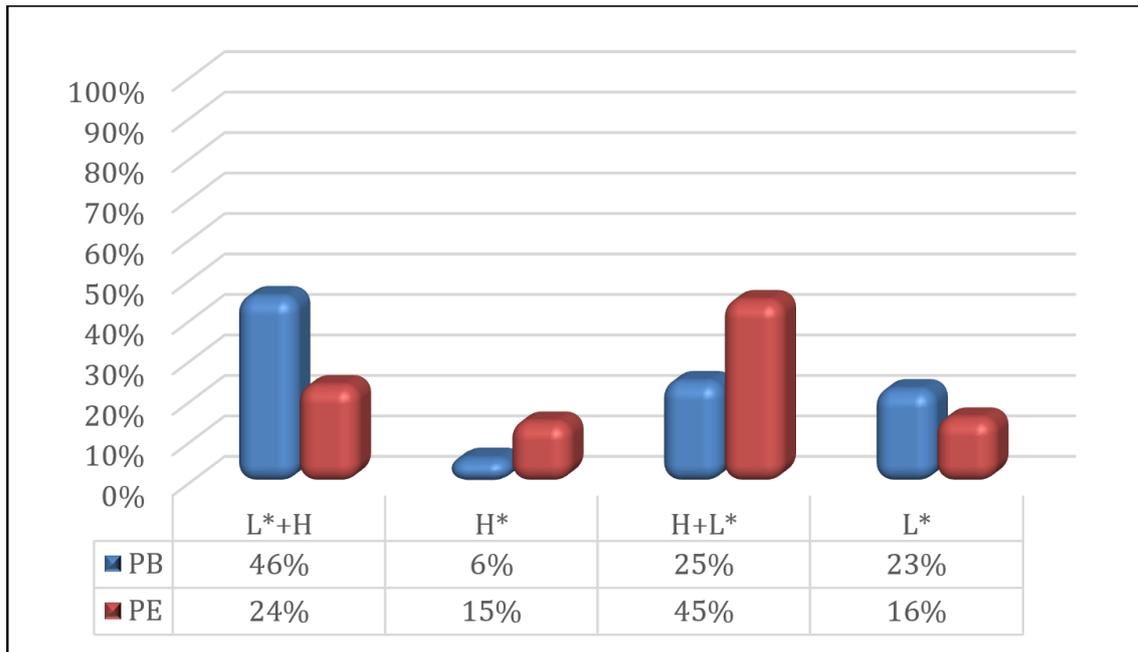
Fonte: Silvestre (2017).

Gráfico 1. Contornos predominantes no PhP inicial orações *não desgarradas* em PB e em PE



Fonte: Silvestre (2017).

Gráfico 2. Contornos predominantes no PhP inicial de orações *desgarradas* em PB e em PE

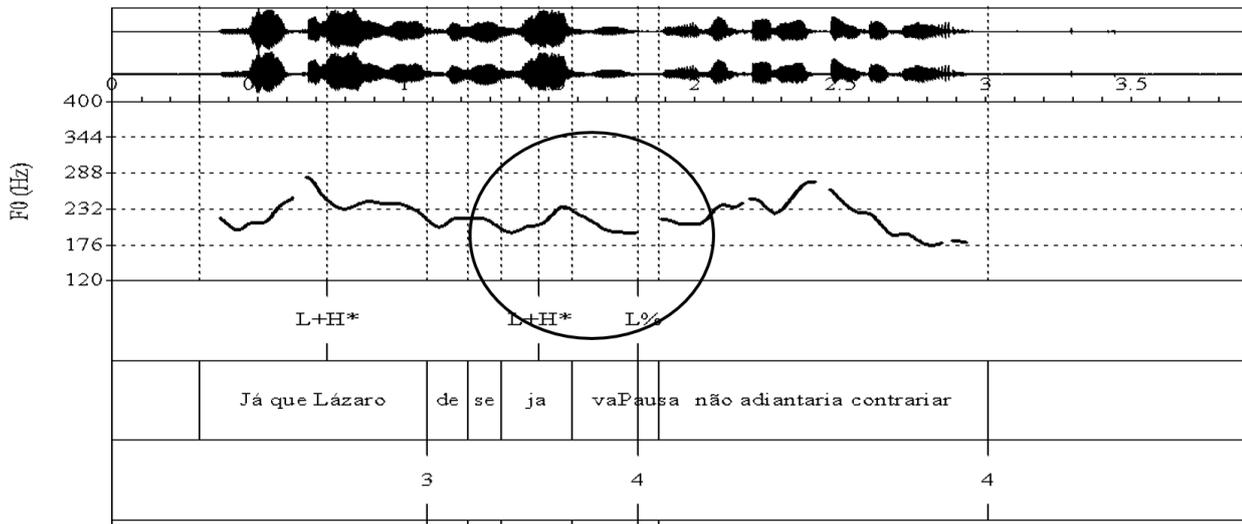


Fonte: Silvestre (2017).

No que tange aos contornos melódicos existentes no fim dos IPs, há, entre orações *não desgarradas* e orações *desgarradas totais*, comportamentos diversos que se mostram capazes de contribuir para a caracterização do fenômeno em estudo, além de diferentes contornos serem observados para o PB e para o PE.

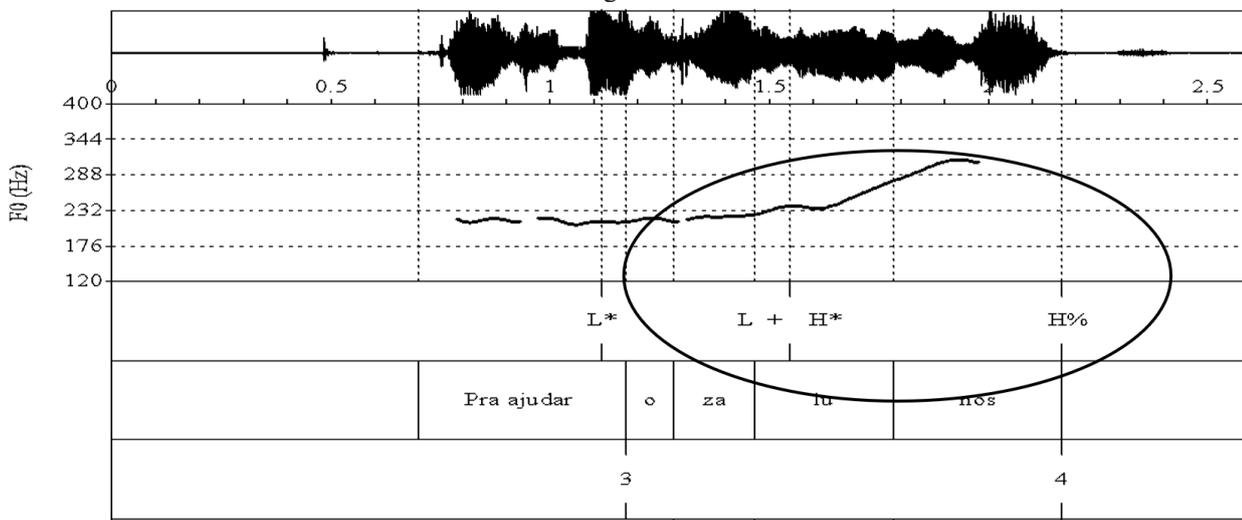
No PB, há perceptível diferença entre orações *não desgarradas* e orações *desgarradas totais*, diferença essa relativa, principalmente, ao tipo de fronteira preferida na produção de cada tipo oracional: predominam os contornos com fronteira baixa – 72% de L% – nos dados sem *desgarramento* ao passo que, nas orações *desgarradas totais*, a predominância de contornos melódicos em que não há descida final é semicategórica (94%) – 83% de L+H*H% e 11% de H+L*LH%. No PE, predominam os tons com fronteira alta ou ascendente – H% ou LH% – tanto nas orações *não desgarradas* quanto nas orações *desgarradas totais*. A figuras de 6 a 9 e os gráficos 3 e 4, a seguir, revelam nossos achados.

Figura 6. Contorno L+H*L% predominante no fim de IPs representativos de orações não desgarradas do PB



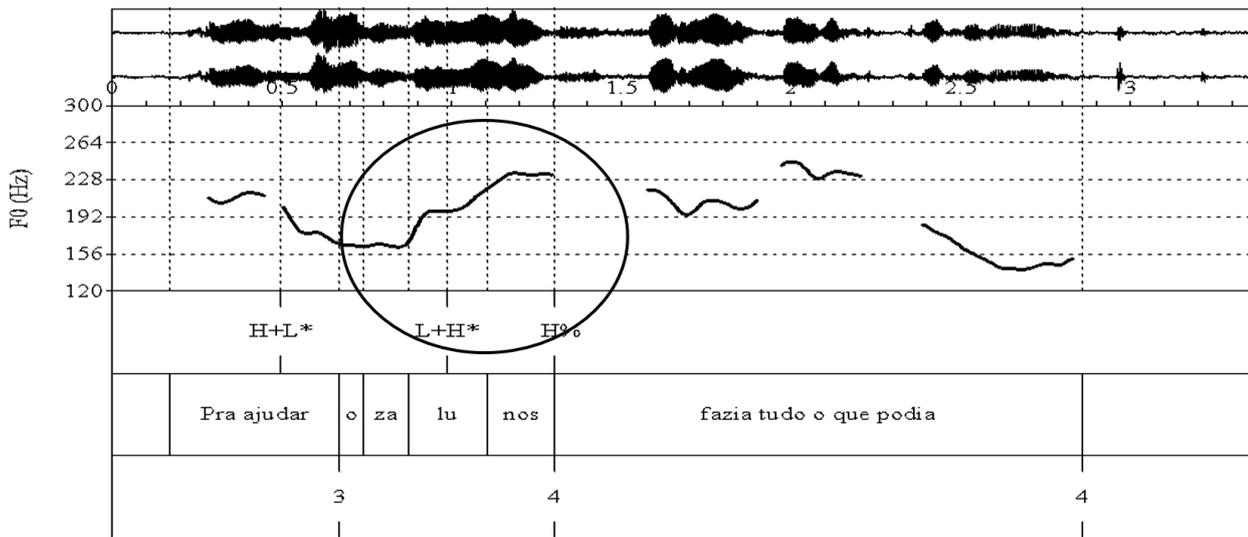
Fonte: Silvestre (2017).

Figura 7. Contorno L+H*H% predominante no fim de IPs representativos de orações desgarradas do PB



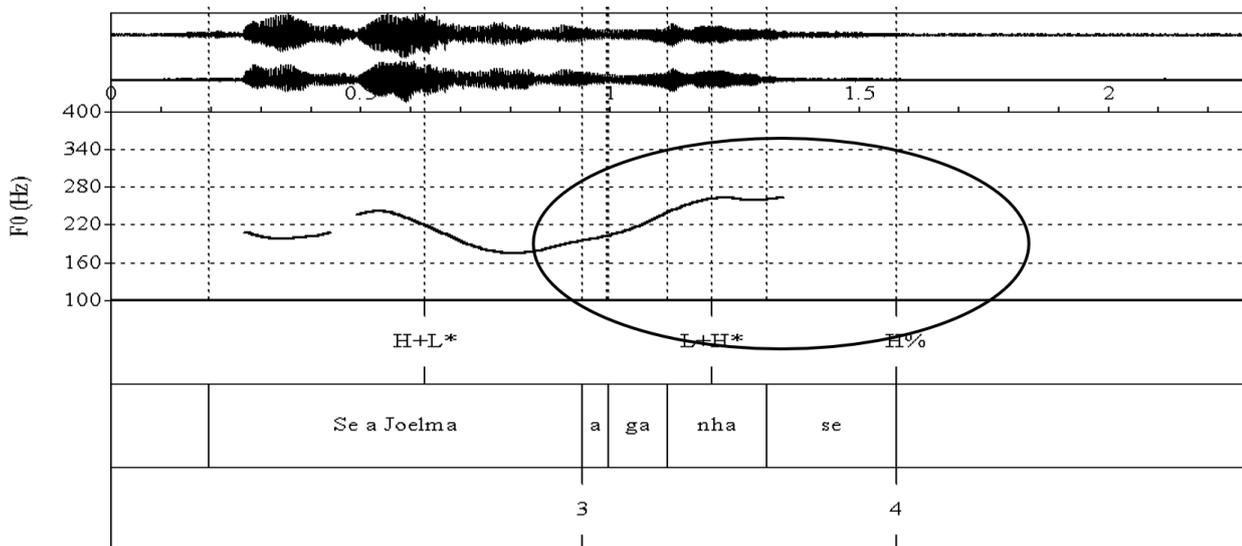
Fonte: Silvestre (2017).

Figura 8. Contorno L+H*H% predominante no fim de IPs representativos de orações não desgarradas do PE



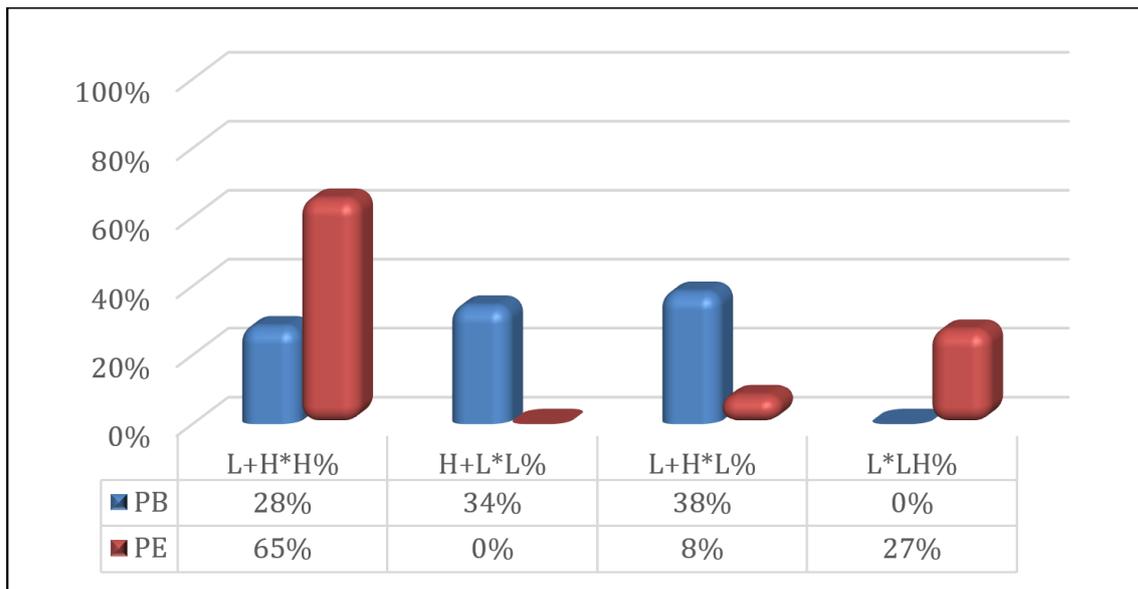
Fonte: Silvestre (2017).

Figura 9. Contorno L+H*H% predominante no fim de IPs representativos de orações desgarradas do PE



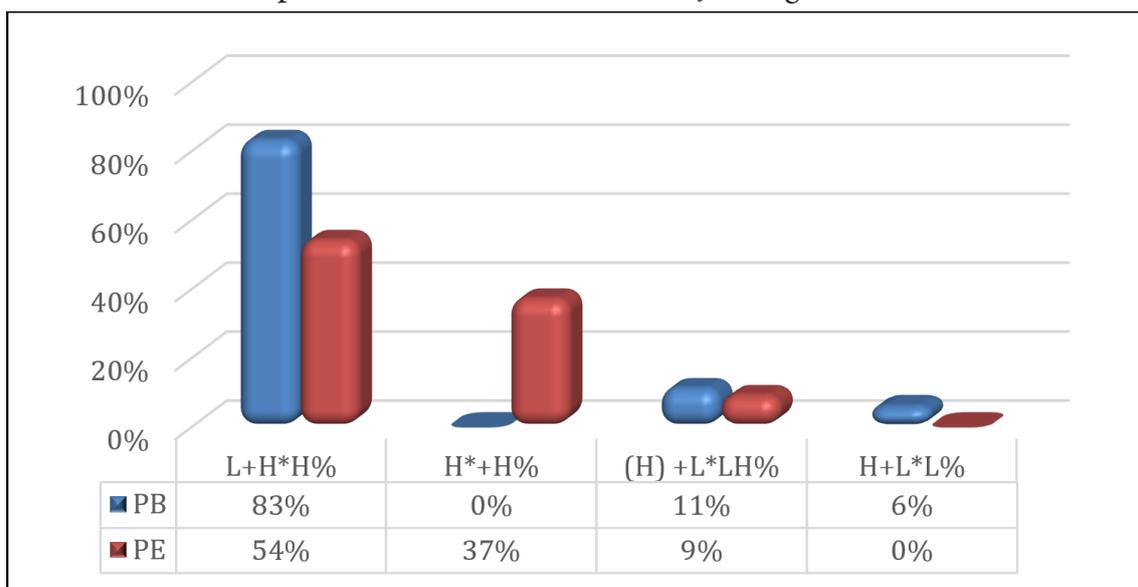
Fonte: Silvestre (2017).

Gráfico 3. Contornos predominantes no PhP final de orações *não desgarradas* em PB e em PE



Fonte: Silvestre (2017).

Gráfico 4. Contornos predominantes no PhP final de orações *desgarradas totais* em PB e em PE



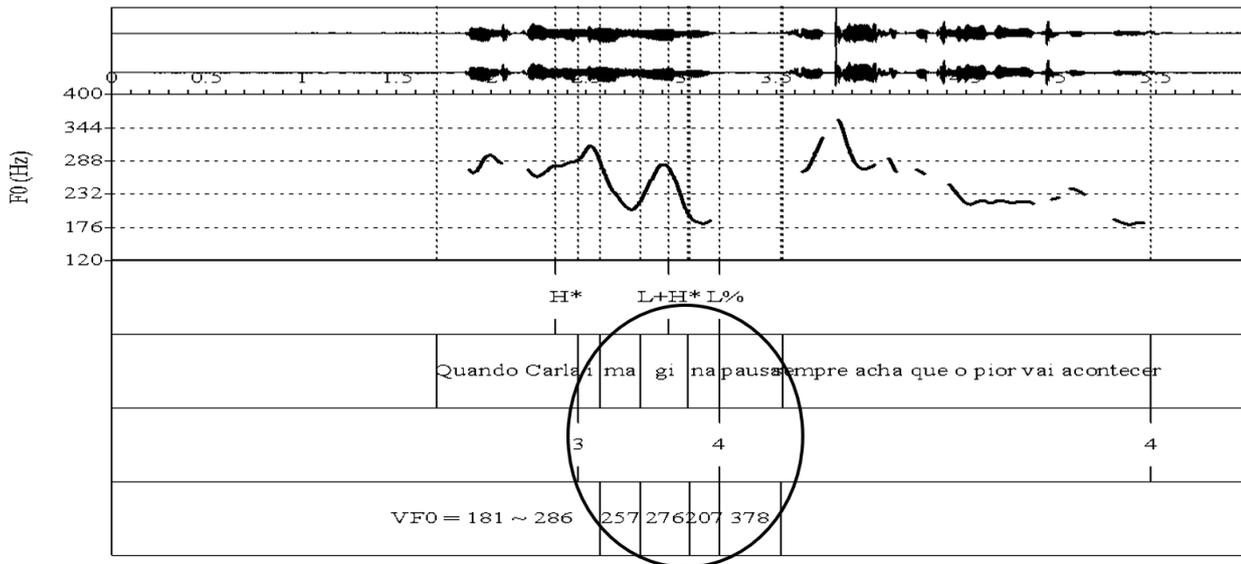
Fonte: Silvestre (2017).

O comportamento duracional das últimas sílabas do IP também mostra resultados diversos para orações *não desgarradas* e orações *desgarradas totais*, tanto em PB quanto em PE, revelando-se, juntamente com os padrões melódicos finais, fator importante na caracterização do fenômeno aqui estudado.

Há, para ambas as variedades, maior duração das sílabas finais nos dados de *desgarra-mento* quando comparadas às mesmas sílabas em orações *não desgarradas*. Em PE, o alongamento médio da sílaba final em relação à última pré-tônica é consideravelmente nas *desgarradas* chega a 26%, ao passo que, nas orações não desgarradas, não ultrapassa 2%.

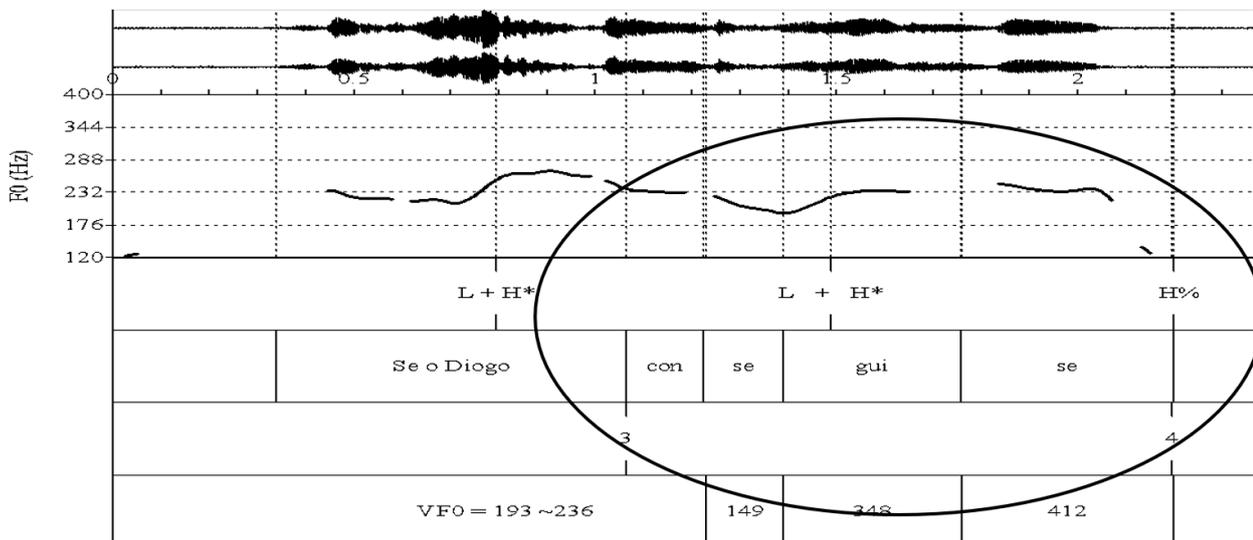
Nas *desgarradas totais* produzidas por falantes brasileiros, acresce-se à maior saliência no alongamento da última pós-tônica (40%, em média) em relação à pré-tônica, também verificada em PE, algum alongamento da sílaba final em relação à tônica, revelando a robustez de tal pista prosódica na caracterização do *desgarramento* em PB. Novamente, as figuras (de 10 a 13) e os Gráficos 5 e 6 exemplificam os resultados:

Figura 10. Duração das sílabas finas dos IPs representativos de orações não desgarradas do PB



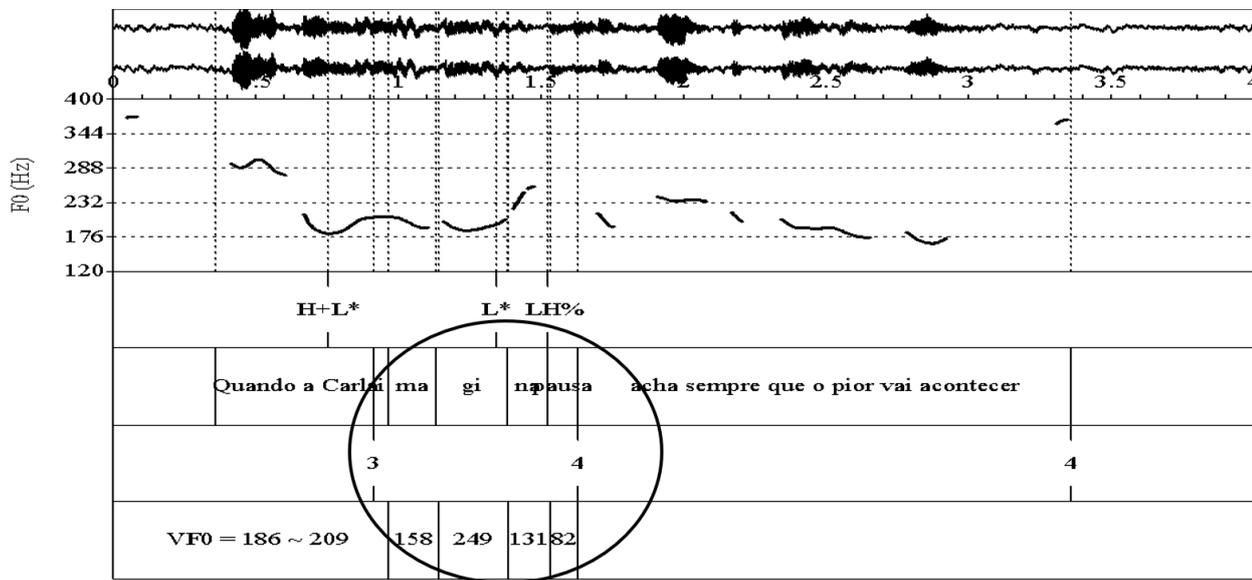
Fonte: Silvestre (2017).

Figura 11. Duração das sílabas finas dos IPs representativos de orações desgarradas do PB



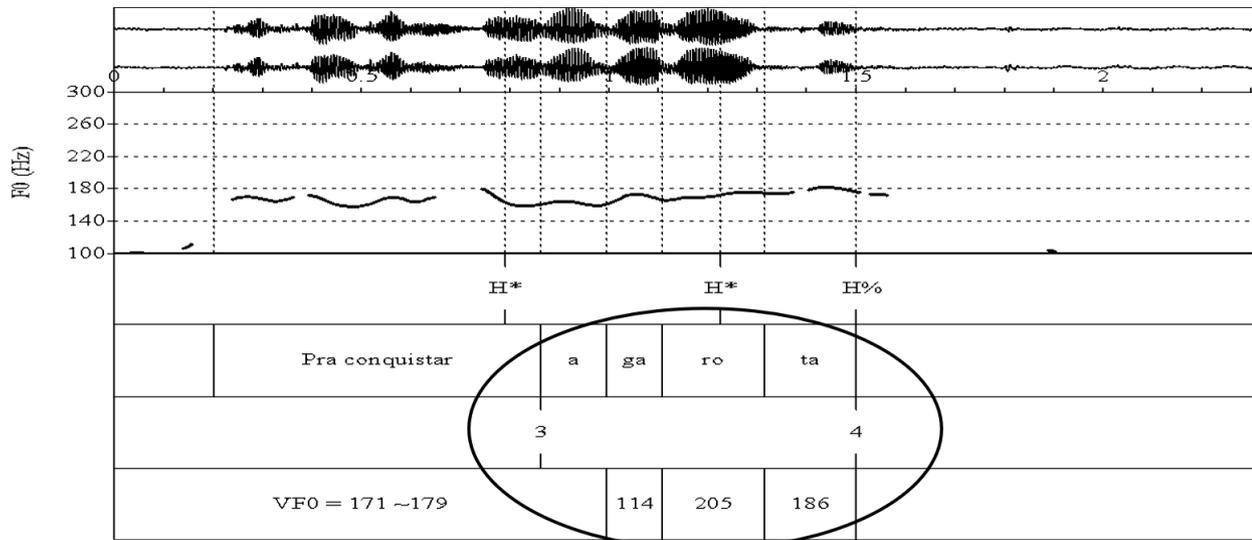
Fonte: Silvestre (2017).

Figura 12. Duração das sílabas finas dos IPs representativos de orações não desgarradas do PE



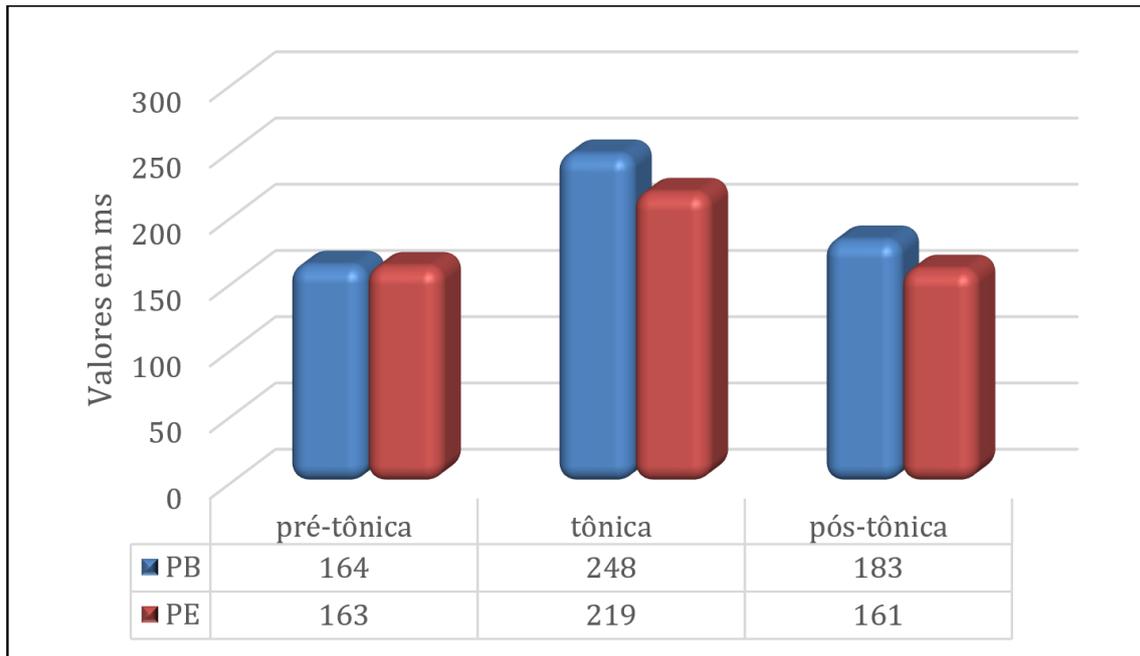
Fonte: Silvestre (2017).

Figura 13. Duração das sílabas finas dos IPs representativos de orações desgarradas do PE



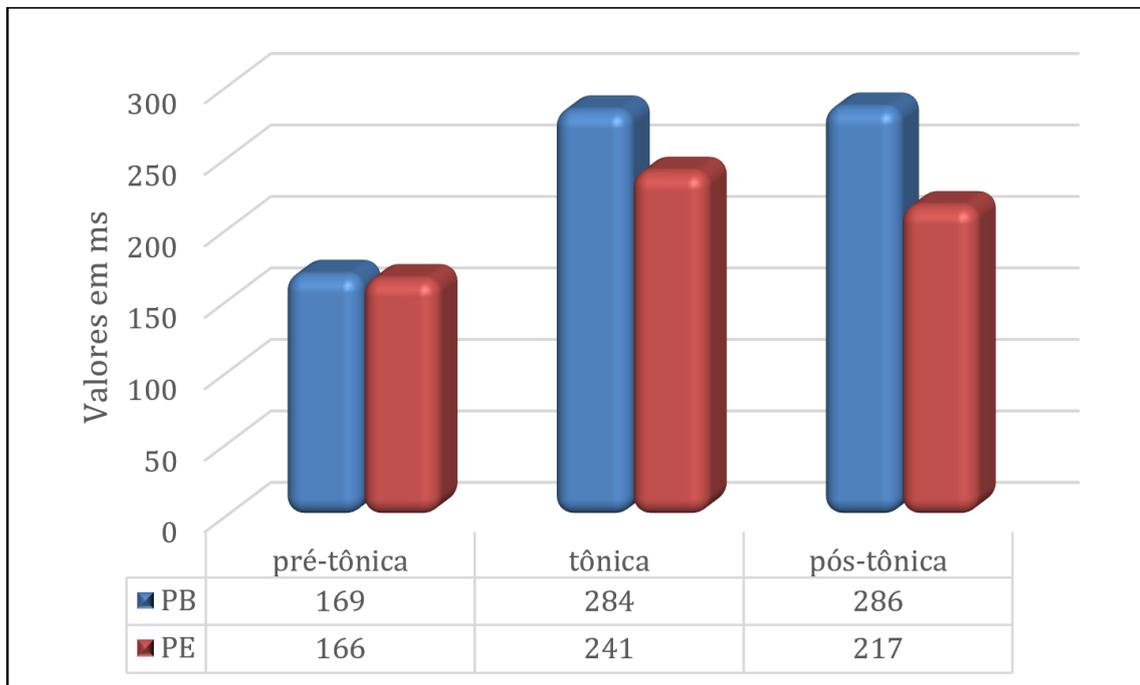
Fonte: Silvestre (2017).

Gráfico 5. Média da duração nas sílabas finais de orações *não desgarradas* em PB e em PE



Fonte: Silvestre (2017).

Gráfico 6. Média da duração nas sílabas finais de orações *desgarradas* em PB e em PE



Fonte: Silvestre (2017).

Os resultados relativos aos contornos melódicos predominantes no fim dos IPs e ao alongamento silábico identificados nas sílabas finais revelam que a interpretação de orações *desgarradas totais* como orações completas se dá pela utilização diferenciada de pistas prosódicas, assim como já demonstraram estudos de Vigário (2003) e Fonseca (2010), por exemplo, para estruturas de desambiguação. No PB, a inserção de pistas prosódicas diferenciadoras se dá de forma ainda mais evidente, o que vai ao encontro de outras descrições comparativas entre PB e PE, como as de Frota e Vigário (2000) e Fernandes (2007), que revelam haver, para o PB, maior saliência de pistas prosódicas pelo fato de ser o PhP ou a PW domínios entoacionais mais robustos na variedade brasileira.

Importa notar que o contorno L+H*H%, majoritariamente presente nas orações *desgarradas* que analisamos, é comumente descrito como característico de um padrão “continuativo” (CAGLIARI 1982, CUNHA 2000, TENANI 2002). Tal fato dá pistas sobre a importância de tal contorno ser acompanhado de outra pista prosódica quando presente na configuração de estruturas de *desgarramento*, já que não há, para as orações *desgarradas totais*, necessidade de continuação sintática ou fonológica. Logo, nossos resultados demonstram que o comportamento diferenciado da duração atua de forma produtiva na concretização do fenômeno em estudo: no PB, salientando ainda mais as diferenças já existentes, uma vez que os contornos predominantes no fim dos IPs já eram diversos nos dados sem ou com *desgarramento*; no PE, estabelecendo que, de fato, as *desgarradas totais* têm comportamento prosódico diferenciado, já que os contornos entoacionais predominantes foram semelhantes no fim de todos os IPs analisados.

Considerações finais

Nossa análise do *desgarramento* mostrou que, tanto em PB quanto em PE, há pistas prosódicas que atuam na diferenciação de orações *desgarradas totais* e que o comportamento da duração nas sílabas finais é elemento comum e importante na verbalização de tais orações. No PB, entretanto, além da perceptível variação fonética estabelecida pelo alongamento, as orações *desgarradas* possuem um padrão melódico final diferente das *não desgarradas*, o que nos permite afirmar que, diferentemente do PE, em que a caracterização do *desgarramento* é dada somente pela pista fonética da duração, em PB as orações *desgarradas* constituem um padrão fonológico diferente.

Na busca de respostas sobre a relação entre prosódia e *desgarramento* no PB e no PE, acabamos, também, por proceder à descrição prosódica das orações adverbiais que são formalmente anexadas à oração núcleo, descrição essa que, com mencionamos revela

dados interessantes no que se refere ao fraseamento prosódico do português, uma vez que, apesar de a literatura da área cunhar o contorno melódico L+H*H% como caracterizador do padrão “continuativo”, sendo a fronteira alta o principal indicador de tal continuidade, as orações *não desgarradas* do PB, que necessitavam de continuação na produção, apresentaram o contorno melódico L+H*L%.¹

Além disso, nossos resultados também contribuem para o fortalecimento da teoria fonológica no sentido de reafirmar que, nos domínios mais altos da hierarquia, a interface entre fonologia e sintaxe é fortemente restrita, não havendo isomorfismo entre os dois componentes da gramática, já que as *desgarradas*, sem lugar na tradição dos estudos sintáticos, são claramente um constituinte fonológico completo, com características prosódicas próprias, não importando, inclusive, a classificação semântica tradicionalmente atribuída à análise sintática de orações adverbiais. Podemos dizer, assim, que, tal qual às estruturas desambigüizáveis referidas por Nespor e Vogel (1994, p. 293), uma oração adverbial sem a chamada oração “principal” existe porque corresponde a uma diferente estrutura prosódica.

Referências

BARROS, N. *Fraseamento prosódico em português: uma análise entoacional de enunciados com parentéticas e tópicos em duas variedades do Português Europeu*. Dissertação de Mestrado. Universidade de Lisboa, Lisboa, 2014.

BECKMAN, M.; HIRSCHBERG, J.; SHATTUCK-HUFNAGEL, S. The original ToBI system and the evolution of the ToBI framework. In: JUN, S.-A. (Ed.). *Prosodic Typology - the Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 2005.

BECKMAN, M. E.; PIERREHUMBERT, J. B. Intonational structure in Japanese and English. *Phonology*, v. 3, p. 255-309, 1986.

BOESMA, P.; WEENICK, D. Praat: doing phonetics by computer [Computer program]. Versão 5.4.08, 2015. Disponível em: www.praat.org.

CASTELO, J. Entoação dos enunciados declarativos e interrogativas no português do Brasil: uma análise fonológica ao longo da costa atlântica. Tese de Doutorado. Universidade de Lisboa, Lisboa, 2016.

CHAFE, WALLACE L. The deployment of consciousness in the production of a narrative. In: CHAFE, W. L. (Ed.) *The Pear Stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood: Ablex, 1980.

CAGLIARI, L. C. *Aspectos acústicos da entoação do português brasileiro*. Série Estudos, n.8, p. 45-59, 1982.

¹ Tais resultados são reforçados por teste perceptivo preliminar realizado por Silvestre (2017).

CUNHA, C. S. *Entoação regional no português do Brasil*. Tese de doutoramento em língua portuguesa. Rio de Janeiro, Faculdade de Letras, UFRJ, 2000.

DECAT, M. B. N. Por uma abordagem da (in) dependência de cláusulas à luz da noção de “unidade informacional”. *Scripta (Linguística e Filologia)*, v. 2, n. 4, p. 23-38, 1999.

_____. *Estruturas Desgarradas em Língua Portuguesa*. Campinas: Pontes Editora, 2011.

FERNANDES, F. R. *Ordem, focalização e preenchimento em português: sintaxe e prosódia*. Tese de Doutorado em Linguística. Campinas: LEL/UNICAMP, 2007.

FONSECA, A. A. *Pistas Prosódicas e o Processamento de sentenças ambíguas do tipo “SN1-V-SN2-Atributo” do Português Brasileiro*. Dissertação de mestrado. Belo Horizonte: UFMG, 2008.

_____. O efeito do peso dos constituintes prosódicos na desambiguação de orações relativas reduzidas. *ReVEL*, v. 8, n. 15, p. 242-255, 2010.

FROTA, S. *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing, 2000.

_____. The intonational phonology of European Portuguese. In JUN, S.-A. (Ed.). *Prosodic Typology II*. Oxford: Oxford University Press, p. 6-42, 2014.

FROTA, S.; VIGÁRIO, M. Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In: CASTRO, R. V.; BARBOSA, P. (Ed.). *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*, v. 1. Coimbra: APL, p. 533-555, 2000.

FROTA, S. et al. Intonational variation in Portuguese: European and Brazilian varieties. In: FROTA, S.; PRIETO, P. (Ed.). *Intonation in Romance*. Oxford: Oxford University Press, 2015. p. 235-283.

HAYES, B.; LAHIRI, A. Bengali intonational phonology. *Natural Language & Linguistic Theory*, v. 9, n. 1, p. 47-96, 1991.

LADD, R. *Intonational phonology*. Cambridge: Cambridge University Press, 2008.

NESPOR, M.; VOGEL, I. prosodic hierarchy na speech perception. *La Perciozone del Linguaggio*, Anais do Seminário, 1980, Florencia, Accademia della Crusca, p. 339-362, 1983a.

_____. Prosodic structure above the word. In: CUTLER, A.; LADD, D. R. (Ed.). *Prosody: Models and Measurements*. Berlim: Springer-Verlag, 1983b. p. 123-140.

_____. *La prosódia*. Tradução de Ana Ardid Gumiel. Madrid: Visor Distribuciones, 1994 [1986].

NEVES, M. H. M. As construções concessivas. In: NEVES, M. H. M. (Org.) *Gramática do português falado*. São Paulo: Humanitas/ FFLCH/USP; Campinas: Editora da Unicamp, v.VII (Novos Estudos). *A gramática: história, teoria e análise*. São Paulo: Editora UNESP, 1999.

_____. A extensão da análise dos elementos adverbiais para além da oração. In: *Revista da ANPOLL*, n. 14, p. 125-137, São Paulo, 2003.

PIERREHUMBERT, J. *The phonology and phonetics of English intonation*. PhD Thesis. Massachusetts: M.I.T, 1980.

PIERREHUMBERT, J; BECKMAN, M. *Japanese tone structure*. Cambridge: M.I.T. Press, 1988.

SELKIRK, E. On prosodic structure and its relation to syntactic structure. Artigo apresentado na Conferência sobre Representação Mental em Fonologia, IULC, 1980.

_____. *Phonology and Syntax: the relation between sound and structure*. Cambridge, MIT Press, 1984.

_____. On derived domains in sentence phonology. In: *Phonology Yearbook* 3. p. 371-405, 1986.

SERRA, C. R. *Realização e percepção de fronteiras prosódicas no português do Brasil: fala espontânea e leitura*. Tese de Doutorado em Língua Portuguesa. Rio de Janeiro: UFRJ/ Faculdade de Letras, 2009.

SEVERINO, C. *Fronteiras prosódicas e desambiguação em Português*. Dissertação de mestrado. Universidade de Lisboa, Lisboa, 2011.

SILVESTRE, A. P. S. A entoação regional dos enunciados assertivos nos falares das capitais brasileiras. Dissertação de mestrado em Língua portuguesa. Rio de Janeiro: UFRJ/ Faculdade de Letras, 2012.

SILVESTRE, A. P. S. *“Se eu pudesse e se o meu dinheiro desse...”: Desgarramento e prosódia no Português Brasileiro e no Português Europeu*. Tese de Doutorado em Língua portuguesa. Rio de Janeiro: UFRJ/ Faculdade de Letras (2017)..

TENANI, L. E. *Domínios prosódicos do português do Brasil: implicações para a prosódia e para a aplicação de processos fonológicos*. Tese de Doutorado em Lingüística. Campinas: LEL/UNICAMP, 2002.

TRUCKENBRODT, H. Phonological phrases: their relation to syntax, focus and prominence. PhD Thesis, M.I.T, 2005.

VIANA, F. S. Towards a P_ToBI. Colaboradores: I. Falé, F. Fernandes, I. Marcarenhas, A. I. Mata, H. Moniz & M. Vigário, 2007. Disponível em: <http://www.fl.ul.pt/dlgr/ SonseMelodias/P-ToBI/P-ToBI.htm>.

VIGÁRIO, M. Prosody and sentence disambiguation in European Portuguese. *Catalan Journal of Linguistics*, v. 2. *Special issue on Romance Intonation*, p. 249-278, 2003. (Edited by Pilar Prieto).



Data de submissão: 15/03/2018

Data de aceite: 20/11/2018

**A MULTI-DIRECIONALIDADE DA TRANSFERÊNCIA
DA DURAÇÃO VOCÁLICA DO PORTUGUÊS COMO L3
PARA A L1 (ESPAÑHOL) E A L2 (INGLÊS):
UM ESTUDO LONGITUDINAL**

MULTI-DIRECTIONALITY IN THE TRANSFER OF VOWEL DURATION
FROM L3 PORTUGUESE TO L1 SPANISH AND L2 ENGLISH:
A LONGITUDINAL STUDY

Letícia Pereyron | [Lattes](#) | leticiapereyron@gmail.com
Escola Superior de Propaganda e Marketing | ESPM-Sul

Ubiratã Kickhöfel Alves | [Lattes](#) | ukalves@gmail.com
Universidade Federal do Rio Grande do Sul | CNPq

Resumo: Este trabalho, à luz da Teoria dos Sistemas Dinâmicos, Adaptativos Complexos (BECKNER et al., 2009; DE BOT et al., 2013; SILVA, 2014), visa a investigar a premissa de que uma mudança em qualquer parte do sistema linguístico pode gerar alteração nas suas outras partes componentes. A partir desta premissa, conduziu-se um estudo de cunho longitudinal. O estudo contou com um aprendiz trilingue, residente no Brasil, falante do espanhol (variedade mexicana) como L1, inglês como L2 e português como L3. Tal estudo contou com instrução formal, de base comunicativa e articulatória, sobre sons vocálicos do Português Brasileiro (L3). Através da instrução formal, visou-se a causar uma modificação acelerada no sistema fonético-fonológico de L3 do aprendiz, para se verificar se tal modificação ocasionaria efeitos sobre a sua L1 (Espanhol) e a sua L2 (Inglês). Verificou-se, para esse fim, a duração das vogais tônicas nas três línguas. A instrução deu-se ao longo de 4 meses, com uma aula de 90 minutos por semana. As coletas contaram com listas de palavras nas três línguas referidas, e ocorreram anteriormente ao período de instrução, durante (a cada 4 semanas) e ao término da instrução formal, de modo a totalizar 5 coletas com o participante. Os achados deste estudo longitudinal evidenciaram que as alterações nas durações das vogais do português (L3), aceleradas pela instrução fornecida, causaram alterações nas durações vocálicas (absoluta e relativa) das demais línguas, devido à interconexão dos sistemas do falante multilingue.

Palavras-chave: Multilinguismo; Transferência linguística; Sistemas Adaptativos Complexos.

Abstract: Following a Dynamic, Complex-Adaptive System account (BECKNER et al., 2009; DE BOT et al., 2013; SILVA, 2014), this study aims to address the assumption that changes in an additional language system (such as the L3) may account for the co-occurrence of changes in the previously learned language systems (such as the L1 and the L2). Departing from this assumption, a longitudinal study was conducted with a Mexican learner (L1 Spanish) who had been residing in Brazil (L2: English; L3: Brazilian Portuguese) at the time of the study. The instruction consisted of weekly 90-minute classes and took place throughout 4 months. The data collections consisted of recordings of word lists in the three languages. The first recording was conducted before the instructional period, the next three recordings were conducted after every four classes, and the fifth recording was conducted at the end of the experiment, totaling 5 recordings. The findings of the longitudinal study provided evidence to the premise that alterations of vowel durations in the L3 system (Brazilian Portuguese), accelerated by formal instruction, may cause alterations in the absolute and relative durations of the vowels of the other languages, due to the interrelation among the language systems of this multilingual speaker.

Key-words: Multilingualism; Language transfer; Complex Adaptive systems.

1 Introdução

O presente trabalho visa a investigar, à luz da Teoria dos Sistemas Dinâmicos, Adaptativos Complexos (BECKNER et al., 2009; DE BOT et al., 2013, SILVA, 2014), a premissa de que uma mudança em um sistema linguístico pode gerar alteração nos demais sistemas linguísticos de um indivíduo, inclusive em línguas já previamente aprendidas. Com base nessa premissa, conduziu-se um estudo de cunho longitudinal que contou com um aprendiz trilingue falante do espanhol (variedade mexicana) como L1, inglês como L2 e português como L3. Durante um período de 4 meses, foi oferecida instrução formal de base comunicativa e articulatória sobre os sons vocálicos [ɛ], como em “peça”, e [ɔ], como em “copo”, presentes na L3, mas ausentes na L1. Partiu-se da hipótese de que a instrução intensiva de tais sons poderia exercer efeitos sobre as durações vocálicas tanto da L1, quanto da L2, devido à completa interconexão entre os sistemas. Dado que as vogais tônicas do português apresentam durações maiores do que no espanhol (SANTOS, 2014; SANTOS; RAUBER, 2016), previu-se que a instrução levaria os aprendizes a produzirem vogais mais longas não somente na L1, mas também em inglês (L2), dado que a duração das vogais mais longas nesta língua também representa um problema para hispânicos (FEIDEN et al., 2016; PEREYRON, 2017).

A metodologia do presente trabalho foi motivada pela premissa de que dados individuais e longitudinais de aprendizagem de línguas adicionais parecem ser mais ricos em termos de informação e de explanação dos processos de desenvolvimento fonético-fonológico na L2, conforme propõem emergentistas (LARSEN-FREEMAN, 1997; 2013; DE BOT; LOWIE; VERPOOR, 2007; LIMA JÚNIOR, 2016a, b). Nesse sentido, o estudo contou com 5 coletas de dados (uma antes, três durante e uma após a instrução), que consistiam na leitura de palavras nas três línguas referidas, inseridas em frases-veículo. O tratamento acústico dos dados deste experimento ocorreu através do programa *Praat* (BOERSMA; WEENINK, 2015). Dessa forma, houve a oportunidade de acompanhar o desenvolvimento individual desse participante trilingue, de modo que fossem traçadas suas curvas desenvolvimentais, conforme sugerem os autores supracitados adeptos à Teoria dos Sistemas Dinâmicos.

Com vistas a atender ao propósito desta pesquisa, o de verificar a premissa de que uma alteração em um sistema pode exercer efeitos sobre os outros, este trabalho está organizado em 5 seções. A seguir, na seção 2, serão apresentados os pressupostos teóricos, seguidos pela Metodologia, que compõe a seção 3. Os resultados, referentes aos valores de duração absoluta e relativa das vogais tônicas nas três línguas, serão apresentados na seção 4, e a conclusão, na seção 5.

2 Pressupostos teóricos

Larsen-Freeman (1997; 2006; 2011; 2013), Ellis (2011) e De Bot et al. (2013), entre outros analisam a língua como um Sistema Adaptativo Complexo (CAS). Nesse sentido, conceber a língua como CAS envolve admitir múltiplos agentes como formadores do próprio sistema linguístico. A complexidade do sistema, entretanto, não resulta dos diversos agentes envolvidos no desenvolvimento linguístico, mas sim das inter-relações que tais múltiplos agentes estabelecem uns com os outros.

Outra propriedade dos Sistemas Adaptativos Complexos é a dinamicidade, o que implica mudança através do tempo. Em sistemas complexos, o resultado do desenvolvimento ao longo do tempo não pode ser calculado com exatidão, não por falta de ferramentas certas, mas porque as variáveis que fazem parte desse sistema interagem e se modificam constantemente. O desenlace dessas variáveis, a menos que fizessem parte de um sistema simples, não pode ser resolvido analiticamente. A trajetória dinâmica, nesse sentido, deve incluir as interações frequentes e recorrentes do próprio sistema e, dessa forma, não há equação que ofereça um valor exato do sistema (DE BOT; LOWIE; VERPOOR, 2007).

De Bot e Larsen-Freeman (2011) afirmam que, em um sistema dinâmico, todas as partes de um dado sistema estão conectadas a todas as outras partes. Conceber a língua como um sistema dinâmico significa considerar que todos os subsistemas - o sistema lexicai, o sistema fonológico e o sistema sintático - estão inter-relacionados e, além disso, uma alteração no sistema fonético-fonológico de uma dada língua poderá resultar em mudanças nas demais línguas do indivíduo, vindo a alterar as demais línguas do falante multilíngue.

Cabe ressaltar, ainda, outra propriedade desses sistemas: a não-linearidade, uma vez que o desenvolvimento linguístico ocorre das interações entre agentes, o que implica resultados imprevisíveis e não-lineares. A não-linearidade também ocorre no âmbito das influências interlinguísticas. A perspectiva que considera o sistema linguístico como CAS postula que as línguas influenciam umas às outras, tanto a mais forte e mais dominante, quanto a mais fraca (FRANCESCHINI, 1999; CENOZ, 2001; BRITO, 2011; GROSJEAN, 2013). A noção de que a transferência está atribuída à influência unidirecional da língua materna, uma vez que aprendizes cometem erros linguísticos que não se referem a suas L1s, é rejeitada. A transferência linguística é, no mínimo, bidirecional.

Considerando a língua, também, como CAS, Herdina e Jessner (2002) aplicam a TSD ao estudo do desenvolvimento multilíngue, de modo a sugerirem que o multilinguismo compõe um sistema dinâmico, composto de sistemas repletos de agentes, que interagem entre si e, assim, geram crescimento. Essas autoras consideram que os sistemas dos falantes bi- ou multilíngues estão inter-relacionados e dependem uns dos outros. Dessa maneira, não é um sistema que influencia o outro de uma maneira linear ou “um para um”; de fato, o sistema linguístico em desenvolvimento é o resultado da inter-relação estabelecida pelos referidos sistemas componentes.

Adicionalmente, Larsen-Freeman (2011), De Bot; Verspoor; Lowie (2007) e Larsen-Freeman; Cameron (2008) e, em contexto brasileiro, Lima Júnior (2016a, b), sugerem uma análise individualizada do desenvolvimento da fala do aprendiz devido à grande dinamicidade e complexidade que compõem o sistema linguístico. Uma vez que um sistema dinâmico complexo e adaptativo é influenciável por uma grande gama de variáveis e é altamente sensível a um estágio inicial, cada indivíduo apresenta, portanto, um sistema de complexidade e singularidades tais que são dignas de ser analisadas individualmente, através da observação do desenvolvimento linguístico do indivíduo ao longo do tempo. Estudos longitudinais e estudos de caso permitem conexões a serem feitas através de níveis e escalas de tempo. Além disso, segundo Barboza (2013), estudos longitudinais

permitem uma descrição do desenvolvimento dinâmico, característica intrínseca aos sistemas adaptativos complexos. O modelo longitudinal apresenta-se como capaz de indicar a variação no desenvolvimento.

Nesse sentido, vem a ser de suma importância um estudo longitudinal, que inclua o papel da interação entre variáveis e, principalmente, que investigue a premissa de que uma mudança em uma das línguas (L3) causa, de fato, alteração nas outras línguas (L1 e L2). Além disso, os modelos dinâmicos apresentam como meta a especificação de como as mudanças dos estados dos sistemas ocorrem. Nesse sentido, modelos dinâmicos têm por objetivo explicar as mudanças temporais. Van Gelder; Port (1995), ao se referirem à premissa de que processos cognitivos se desenvolvem ao longo do tempo real, definem 'tempo' como uma quantidade contínua medida por números reais, e afirmam, ainda, que para cada ponto no tempo há um estado do sistema cognitivo.

Sob essa mesma ótica, a que retoma a noção de 'período', Lima Júnior (2016a, b) analisa o desenvolvimento fonológico da língua adicional com um sistema dinâmico, adaptativo complexo. O autor sugere que a Teoria dos Sistemas Adaptativos Complexos dá conta não apenas de processos linguísticos, mas de sistemas/processos naturais do mundo. Com uma base teórica ancorada em Larsen-Freeman (1997), Herdina; Jessner (2002), De Bot; Lowie; Verspoor (2007), De Bot (2008) e Larsen-Freeman; Cameron (2008), Lima Júnior (2016a, b) propõe uma análise individualizada e longitudinal do desenvolvimento fonológico dos aprendizes, já que, dessa forma, é possível retratar o processo em contínua mudança e não apenas um retrato (estágio) em que se encontra a fala do participante no momento da coleta. Conforme aponta o autor, a análise transversal e aglutinada apresenta-se limitada quando comparada com a análise longitudinal e individual, pela natureza complexa e dinâmica do processo de desenvolvimento de uma LA. Além disso, a complexidade e dinamicidade desse processo está intimamente ligada à não-linearidade, que também é mais bem retratada em estudos individuais e longitudinais.

3 O estudo

3.1 Objetivo, hipótese e operacionalização

O presente estudo teve como objetivo geral verificar a premissa dinâmica de que uma mudança em um componente do sistema linguístico pode modificar os demais componentes. Mais especificamente, o presente estudo longitudinal investiga o papel da intervenção intensiva no desenvolvimento da L3 e seus efeitos sobre os sistemas da L1 e da

L2 do aprendiz trilingue, no que diz respeito aos valores de duração absoluta e relativa das vogais tônicas das palavras investigadas nas três línguas.

A hipótese é a de que a alteração, ao longo do tempo, provinda da instrução da L3 (que leva o aprendiz a abandonar o padrão duracional de sua L1 e passar a produzir vogais mais longas, como no Português) deve modificar a L1 e a L2, devido à completa interconexão entre os sistemas. Dessa forma, a produção das vogais abertas do português [ɛ], como em “peça” e [ɔ], como em “copo”, acelerada pela instrução, poderá causar alterações nos valores de duração tanto absoluta quanto relativa das vogais abertas do inglês, sobretudo [ɛ], como em “set”, [æ], como em “cat”, [ɔ], como em “often” e [ɑ], como em “soccer”, mesmo sendo essas últimas distintas das vogais abertas do português, e também nos valores das durações das vogais [a], como em “paco”, [e], como em “pepa”, e [o], como em “coche” do espanhol. Conforme já explicado, espera-se que a instrução do português possa causar durações mais longas nas vogais da L2 e também da própria L1, dado que as vogais do português tendem a ser mais longas do que as do Espanhol (SANTOS, 2014; SANTOS; RAUBER, 2016).

Quanto à operacionalização, foram comparadas as curvas desenvolvimentais dos sistemas vocálicos das línguas do aprendiz trilingue longitudinalmente: (uma semana) antes do início do período de instrução, ao longo do período de instrução (a cada quatro semanas) e ao fim de todas as sessões de instrução, com vistas a verificar se houve possíveis alterações nas durações vocálicas, em cada um dos sistemas, ao longo de todo o período de investigação¹.

3.2 O informante

O estudo longitudinal (que transcorreu de abril a agosto de 2015) contou com um participante de nacionalidade mexicana, com 38 anos de idade e aluno de Doutorado do Programa de Pós-Graduação em Biologia em uma instituição privada de ensino superior de Porto Alegre. O aluno demonstrou forte interesse e motivação para aprender o português. Seu contato inicial com a língua portuguesa havia começado 5 meses antes da sua vinda ao Brasil, através dos cursos *Duolingo* e *Babel*.² O aprendiz mexicano convive com outros falantes de espanhol na sua residência, mas afirma usar a língua portuguesa em

¹ Para o estudo completo, veja Pereyron (2017), que apresenta, além das durações, as descrições das alterações em termos de altura e anterioridade/posterioridade que as vogais das três línguas do participante sofreram.

² Ambos os cursos são oferecidos na internet, gratuitamente. Sua metodologia insere jogos e outras atividades interativas.

contexto acadêmico, como nas aulas do Doutorado. O participante residia há dois meses em Porto Alegre quando iniciou o presente estudo. Quanto ao inglês, o aprendiz informou que o adquiriu de forma autônoma, lendo artigos e traduzindo músicas. Ademais, possui nível de proficiência B1 na língua inglesa de acordo com o resultado obtido no *Oxford Online Placement Test*³, sendo, portanto, um aprendiz de grau intermediário-avanzado.

3.3 Palavras e Instrumentos de coleta

Para fins da presente pesquisa, foram formulados três instrumentos de leitura: uma sequência de frases-veículo com palavras na língua espanhola, uma sequência com palavras na língua portuguesa e outra, com palavras em inglês. Os vocábulos dos instrumentos em português e em espanhol são dissilábicos (como, por exemplo, ‘faca’ em português, e ‘pepa’, em espanhol) e trissilábicos (como ‘sucata’, em português, e ‘machete’, em espanhol). Optou-se por palavras com duas e três sílabas para atender ao critério de que as palavras tivessem que ser bastante frequentes⁴ na língua a ser lida, visto que a variável referente à frequência de uso foi controlada. Para tal propósito, apenas palavras dissilábicas não foram suficientes; assim, foram inseridas as palavras com três sílabas e também bastante frequentes, como, por exemplo, a palavra ‘alface’.

A vogal analisada foi a tônica, localizada na primeira sílaba da palavra dissilábica (como, por exemplo, a vogal ‘a’ em ‘fato’), ou na segunda sílaba da palavra de três sílabas (como, por exemplo, a vogal ‘a’ em ‘sucata’). Para ambas as línguas foram utilizadas as plosivas [p, t, k] e/ou as fricativas e africadas [s, ʃ, tʃ, f] como contexto antecedente e seguinte. Os contextos antecedentes e seguintes da vogal tônica eram surdos⁵, de modo a não causar um alongamento da referida vogal devido ao vozeamento da consoante precedente ou seguinte (cf. ALVES, 2015). Foram investigadas as cinco vogais do espanhol [a, e, i, o, u] e as sete vogais orais do português [a, ɛ, e, i, o, ɔ, u]. Para cada vogal, foram

³ O Oxford Online Placement Test é um teste online da Oxford University Press, que mede tanto a habilidade de compreensão da gramática da língua inglesa, quanto a compreensão oral. Os testes podem ser comprados em www.oxfordenglishtesting.com. Para maiores detalhes acerca do teste, conferir Purpura (2007).

⁴ Um critério tomado na escolha das palavras em língua inglesa e portuguesa foi o fato de essas palavras não serem cognatas com a língua espanhola. Além disso, a frequência das palavras foi verificada através do banco de dados *corpus brasileiro* do LAEL (Linguística Aplicada e Estudos da Linguagem) da PUCSP, para as palavras em português. Já as palavras em inglês tiveram a frequência verificada no site www.clearpond.com. Quanto às palavras em espanhol, essas foram verificadas através do *Espal – Spanish lexical database*.

⁵ Houve apenas uma exceção com referência ao vozeamento. Trata-se da palavra ‘beco’, em português, que apresenta contexto antecedente vozeado, pois não foi encontrada uma palavra dissilábica com a vogal [e] e contexto antecedente surdo que não fosse cognata.

apresentadas seis palavras (6 *types*), com três repetições (3 *tokens*), de modo a serem obtidas dezoito produções de cada vogal em cada língua. Além disso, não foram incluídas palavras cognatas com o espanhol no instrumento em português, a fim de não induzir o informante falante de espanhol como L1 a acessar sua língua nativa. O mesmo procedimento foi realizado para o espanhol: o instrumento contou com seis palavras para cada vogal, três sendo dissilábicas e três trissilábicas, repedidas três vezes, totalizando dezoito produções, por participante, para cada vogal.

As palavras foram inseridas em frases-veículos (Diga “palavra-alvo”), apresentadas em *slides*, como “diga tosse” e “diga pipoca”, em português, e “diga coche” e “diga machaca”, em espanhol; dessa forma, a taxa de elocução (velocidade de fala) pôde ser controlada. A lista de palavras do espanhol contou com 105 *tokens*, dos quais 90 eram palavras-alvo e 15 distratores; por sua vez, a lista de palavras do português contou com 145 *tokens*, dos quais 126 eram palavras-alvo e 19, distratores.

Para a formulação do instrumento em inglês, foram utilizadas palavras monossilábicas, como em ‘*cat*’, e dissilábicas, como em ‘*apple*’, uma vez que esses padrões são mais frequentes na língua. Foram investigadas as nove vogais do inglês norte-americano, conforme o estudo de Blank (2013): [ɑ, æ, ɛ, ɪ, i, ɔ, ʊ, ʌ, u]. Para cada vogal, também utilizaram-se as plosivas surdas [p, t, k], fricativas e africadas surdas [s, ʃ, tʃ, f], como contextos antecedentes e seguintes. Optou-se por apenas essas obstruintes, e não nasais ou líquidas, para uma análise acústica mais clara, uma vez que essas consoantes se mostram mais visualmente perceptíveis nos espectrogramas. As consoantes vozeadas também foram excluídas como contexto antecedente ou seguinte, por promoverem um alongamento da vogal (ALVES, 2015). Os critérios empregados foram os mesmos dos outros instrumentos; dessa forma, o instrumento em inglês também foi composto por palavras inseridas em frases-veículo (“Say _____”), contou com 6 *types* por vogal, 3 repetições, totalizando 18 *tokens* por vogal. Quanto ao número de palavras do instrumento em inglês, a leitura incluiu 186 *tokens*, dos quais 24 eram distratores. Além disso, todas as palavras foram inseridas, nos respectivos instrumentos, em ordem aleatória.

3.4. As coletas

As coletas foram realizadas em uma sala silenciosa, onde apenas o participante e a primeira autora deste trabalho estavam presentes. Ao participante, foi solicitada a leitura das palavras dos três instrumentos da presente pesquisa. A primeira coleta foi realizada antes do período de instrução e as seguintes coletas ocorreram a cada 4 semanas durante

os meses de instrução, de modo que a quinta coleta ocorreu ao final desse período. As coletas foram realizadas com pausas de 5 minutos entre as leituras dos instrumentos. O ordenamento de aplicação dos instrumentos em espanhol e em português deste estudo foi realizado de forma alternada, a fim de que o participante não se acostumassem com a ordem dos instrumentos de leitura. Além disso, procurou-se iniciar com o instrumento nas línguas que estavam ativadas, ou seja, espanhol e português. Desse modo, a leitura do inglês sempre foi a última a ser realizada, já que não estava sendo utilizada oralmente durante o período de instrução. O *software* utilizado para a gravação de áudio dos participantes foi o *Audacity*⁶, e a gravação foi realizada em um computador *Sony Vaio*, com um microfone *Sony*, modelo ECM XM1, ligado diretamente no computador. Todo o material acústico foi salvo em arquivo *wav*, com taxa de amostragem de 44.100 Hertz.

3.5 A instrução

A instrução, com base no ensino de pronúncia de cunho comunicativo (cf. CELCE-MURCIA et al., 1996), além de voltada ao desenvolvimento de aspectos de pronúncia não em contexto isolado, mas integrado a funções comunicativas, teve por objetivo diminuir as dificuldades dos falantes nativos do espanhol na produção vocálica em português, como, por exemplo, na produção das vogais médias baixas. Teve, também, o objetivo de acelerar o processo de desenvolvimento fonético-fonológico do participante, para que, no curto intervalo de quatro meses, pudessem ser vistas modificações na produção das vogais da L3 do aprendiz. Desse modo, ao se enfatizarem os sons [ɛ] e [ɔ] inexistentes na língua espanhola, mas existentes na portuguesa, pretendeu-se forçar mudanças no sistema de L3 e verificar se tais mudanças alterariam os outros sistemas de L1 e L2. As sessões foram ministradas em 16 encontros com duração de uma hora e 30 minutos cada, sendo que as 8 primeiras sessões ocorreram no curso de português para estrangeiros (com foco em escrita e leitura), juntamente a outros alunos que faziam parte da instrução. Em seguida, por mais 8 sessões, o participante mexicano participou da instrução de forma individual, cujas aulas aconteceram na biblioteca da instituição.

A seguir, serão apresentados os resultados, em termos de duração absoluta e relativa⁷, referentes às vogais da língua de instrução, o português, para a verificação de uma possível alteração em termos de duração das vogais do espaço acústico da L3 do informante. Em seguida, serão apresentados os resultados referentes às durações das vogais da

⁶ *Software* livre. Download gratuito em www.audacity.sourceforge.net.

⁷ Em Pereyron (2017), encontra-se uma descrição completa do estudo em questão, em que os valores formânticos também são apresentados.

L2 do aprendiz, o inglês, para uma verificação de possíveis alterações do sistema vocálico, sobretudo das vogais mais baixas, dessa língua. Finalmente, serão apresentados os resultados referentes à L1 do aprendiz, o espanhol, com a finalidade de verificação da premissa que postula que uma alteração em um dos sistemas, no presente caso, o português, possa vir a alterar todo o restante, inclusive o sistema materno.

4 Resultados

4.1 O português

A instrução intensiva, juntamente à imersão em contexto de L3, ocasionou uma produção mais abaixada das vogais médias-baixas anteriores da L3 do aprendiz a partir da terceira coleta, conforme é demonstrado em Pereyron (2017), no que diz respeito aos formantes. A produção (e a percepção) das vogais médias anteriores de forma mais baixa vão ao encontro do que afirma Lindblom (1986), com relação à melhor percepção das vogais anteriores do que as vogais posteriores.

Nesse sentido, a realização mais aberta dessas vogais tende a condicionar um alongamento justamente devido à maior abertura da mandíbula para a realização. No entanto, todo o sistema vocálico demonstra maior duração a partir da terceira coleta. A Tabela 1 a seguir apresenta os valores das médias da duração vocálica absoluta do participante do estudo longitudinal, ao longo do período de instrução. Os valores de durações relativas são apresentados na Tabela 2. As células em amarelo correspondem às realizações mais longas pelo participante.

Tabela 1. Médias (em milissegundos) e Desvios-Padrão das durações absolutas das vogais do português produzidas pelo participante do estudo longitudinal (n. de cada vogal = 18 por coleta, total de 126 vogais por coleta)

	[a]	[ɛ]	[e]	[i]	[ɔ]	[o]	[u] ⁸
Coleta 1	142,9 DP:13,5	129,2 DP:16,95	125,7 DP:20,4	97,2 DP:22,8	135,3 DP:14,3	135,5 DP:13,0	117,2 DP:13,0
Coleta 2	167,8 DP:33,4	163,05 DP:26,95	152,5 DP:152,55	113,9 DP:20,0	152,8 DP:28,05	144,2 DP:20,8	126,2 DP:26,2
Coleta 3	184,8 DP:18,03	182,9 DP:16,85	169,4 DP:22,5	126,9 DP:20,4	167,9 DP:27,8	164,1 DP:17,3	128,05 DP:19,3
Coleta 4	168,8 DP:21,0	171,8 DP:15,3	168,2 DP:21,05	112,7 DP:24,7	157,3 DP:20,0	156,2 DP:19,4	128,4 DP:21,9
Coleta 5	165 DP:18,0	165,9 DP:17,8	160,4 DP:20,3	103,6 DP:32,4	153,5 DP:22,4	146,7 DP:15,2	119,8 DP:21,15

Tabela 2. Médias (em porcentagens) e Desvios-Padrão das durações relativas⁸ das vogais do português produzidas pelo participante do estudo longitudinal (n. de cada vogal = 18 por coleta, total de 126 vogais por coleta)

Dissílabos	[a]	[ɛ]	[e]	[i]	[ɔ]	[o]	[u]
Coleta 1	15,88 DP:1,04	13,96 DP:1,63	16,09 DP:2,5	11,31 DP:1,4	15,10 DP:1,5	14,36 DP:1,22	12,32 DP:0,95
Coleta 2	16,17 DP:2,7	15,56 DP:2,06	18,85 DP:2,87	12,20 DP:1,68	16,11 DP:3,0	13,98 DP:1,49	13,53 DP:1,65
Coleta 3	18,92 DP:1,42	18,55 DP:1,72	20,30 DP:2,75	13,75 DP:2,41	18,19 DP:2,6	17,77 DP:2,42	13,21 DP:1,84
Coleta 4	16,49 DP:2,66	16,48 DP:1,4	18,67 DP:2,26	13,20 DP:1,36	16,83 DP:2,16	17,02 DP:1,67	12,22 DP:2,15
Coleta 5	16,50 DP:1,23	17,03 DP:1,45	19,28 DP:2,16	10,85 DP:2,15	17,05 DP:2,16	16,38 DP:1,67	12,36 DP:1,33
Trissílabos	[a]	[ɛ]	[e]	[i]	[ɔ]	[o]	[u]
Coleta 1	14,27 DP:1,13	13,24 DP:1,8	11,61 DP:1,3	9,01 DP:2,53	12,85 DP:1,15	13,31 DP:1,4	11,87 DP:1,07
Coleta 2	16,67 DP:2,16	15,30 DP:2,22	12,53 DP:1,5	9,93 DP:1,8	13,44 DP:1,56	13,04 DP:1,93	11,72 DP:2,5
Coleta 3	17,62 DP:0,92	17,57 DP:1,74	14,62 DP:1,45	11,54 DP:1,35	14,70 DP:1,65	14,66 DP:1,27	12,62 DP:2,1
Coleta 4	15,40 DP:2,24	15,35 DP:1,93	14,26 DP:1,8	8,58 DP:2,33	12,66 DP:0,76	12,92 DP:1,85	11,71 DP:1,42
Coleta 5	15,58 DP:1,7	15,91 DP:1,9	13,85 DP:2,0	9,58 DP:3,7	13,35 DP:1,7	13,39 DP:1,63	12 DP:2,54

Os valores mais baixos (ou seja, as durações vocálicas mais curtas) são sempre encontrados na primeira coleta do participante deste estudo, o que representa o período pré-instrucional, de modo a indicar que o sistema vocálico ainda se encontrava no início do desenvolvimento, visto que o participante já tinha contato com o português neste período. Embora esses valores sejam os mais baixos, a duração vocálica do aprendiz já pode estar mostrando alterações nessa fase, devido à exposição ao sistema alvo, se comparada à duração do sistema vocálico materno do participante, que dispõe de 5 vogais dentre as 7 vogais do português.

Os valores da segunda, quarta e quinta coletas mostram aproximação entre si, mas se distinguem da terceira coleta, que apresenta valores bastante altos, e da primeira, que apresenta os valores mais baixos, o que traz uma noção de não-linearidade ao desenvolvimento do sistema, característica intrínseca da visão de língua como CAS.

É na terceira coleta, no entanto, que se podem observar os valores mais altos de duração absoluta do sistema vocálico do participante deste estudo, representado pelas células em amarelo. Esses valores, ademais, mostram uma aproximação aos obtidos pelos

⁸ Para a obtenção da duração relativa, foi necessário a distinção entre palavras dissilábicas e trissilábicas.

monolíngues porto-alegrenses, apresentados em Pereyron (2017), fato que indica que o sistema do aprendiz está em pleno desenvolvimento em direção às durações do sistema alvo, além de evidências de que a instrução na L3 tenha sido eficaz. Os valores de duração absoluta do participante do estudo longitudinal, na terceira coleta, se aproximam consideravelmente dos de duração absoluta dos monolíngues porto-alegrenses, sobretudo no caso das vogais [a] (198,12 ms), [ɛ] (195,2 ms), [e] (175,13 ms), [i] (145,6 ms), [o] (172,23 ms) e [u] (146,9 ms). A duração absoluta da vogal [ɔ] do participante não parece ainda se assemelhar à dos monolíngues porto-alegrenses, que apresentam valor de 204,9 ms, justamente a vogal mais longa do par [o] e [ɔ], cujo contraste em termos de F1 e F2 não parece ter sido adquirido pelo participante do estudo longitudinal⁹. Esses resultados, referentes a terceira coleta, fortalecem a evidência de que esse período foi marcado por um momento de mudança no sistema por parte do informante, tanto em termos de formantes quanto em termos de duração, de modo a demonstrar não somente a distinção entre as alturas das vogais médias anteriores, como também um alongamento em todas as vogais do português. Além disso, ainda que nas quarta e quinta coletas os valores referentes à duração absoluta sejam mais baixos do que aqueles encontrados na terceira coleta, eles tendem a ser mais altos do que aqueles encontrados no período pré-instrucional, sobretudo no caso das vogais médias-baixas.

Esses achados sugerem que o aprendiz mexicano (que emprega durações quase tão altas quanto às dos nativos brasileiros já na terceira coleta) parece empregar valores de durações intermediárias, possivelmente devido às interações entre os três sistemas do aprendiz. As durações absolutas do aprendiz, sobretudo na terceira coleta, parecem carregar padrões de duração do sistema de língua inglesa, que também dispõe de vogais longas, como também da própria L1, que dispõe de vogais mais curtas do que o português (SANTOS, 2014; SANTOS; RAUBER, 2016). Tal ocorrência se encontra em plena consistência com a ótica de língua como CAS, que prevê a emergência através das interações de estruturas linguísticas, e não a soma delas. Além disso, a ascensão gradiente dos valores da primeira coleta à terceira, bem como o recuo a partir de tal coleta, confirmam o caráter não-linear dos sistemas adaptativos complexos. Cabe por ora ressaltar que, apesar desse recuo, os valores da quinta coleta são maiores do que os da primeira coleta.

Em suma, as durações do sistema vocálico do português do participante deste estudo demonstram uma curva inicialmente ascendente, isto é, os valores aumentam conforme a exposição/imersão do sistema alvo e ao período de instrução formal com referência às vogais do português, que são mais longas do que as do sistema materno, conforme já mencionado, e atingem o pico mais alto na terceira coleta. Após este momento, os valores

⁹ Conforme mostra Pereyron (2017).

mostraram uma redução quanto à duração vocálica, um momento marcado por grande instabilidade.

Assim como os valores de duração absoluta, os valores referentes à duração relativa da terceira coleta se apresentam como os mais altos, de modo a se assemelhar aos valores das durações relativas dos monolíngues porto-alegrenses (PEREYRON, 2017). Apesar de que as durações absolutas e relativas tendam a expressar as mesmas tendências, as relativas parecem fornecer mais evidências para o fato de que a duração vocálica empregada pelo participante mexicano se encontra muito próxima às porcentagens dos nativos porto-alegrenses.

Há uma semelhança nas curvas desenvolvimentais do sistema do português com o do inglês do referido participante (cf. seção a seguir), o que corrobora fortemente a premissa de interconexão e inter-relação entre os sistemas, conforme sugerem os adeptos à TSD e à Teoria da Complexidade (LARSEN-FREEMAN, 1997, 2011; JESSNER, 2008; DE BOT; LOWIE; VERSPOOR, 2007; ELLIS, 2011; LIMA JÚNIOR, 2016a, b; BLANK, 2016). A seguir, será descrita a próxima língua do participante deste estudo, o inglês (L2), que também será verificada em termos de durações absolutas e relativas. A partir da premissa expressa acima, espera-se que, pela instrução das vogais abertas no português (L3), as vogais abertas do sistema de L2 sofram alterações devido à completa interconexão dos sistemas fonético-fonológicos deste falante multilíngue.

4.2 O inglês

O aluno deste estudo também é falante de inglês e possui nível B1 de proficiência em língua inglesa, comprovado pelo *Oxford Online Placement Test*¹⁰, realizado para fins deste estudo. O aprendiz relatou nunca ter cursado aulas de instrução formal de inglês, apenas algumas disciplinas da língua na escola, mas que seu primeiro contato com o idioma foi ao traduzir músicas para o espanhol, alguns anos antes deste estudo. Já na universidade, o aprendiz obteve bastante prática com a língua escrita, através da leitura de artigos acadêmicos.

A interconexão dos sistemas linguísticos do participante deste estudo, característica dos sistemas dinâmicos adaptativos complexos, parece ter agido sobre os valores de duração no sistema de L2 do aprendiz, assim como agiu no sistema de L3, o português (cf. descrito na seção anterior). Grosso modo, as vogais da L3, que apresentaram as durações mais longas na terceira coleta, são justamente as vogais que correspondem às da L2, que também foram alongadas na terceira coleta. As Tabelas 3 e 4, a seguir, apresentam as Médias e Desvios-Padrão dos valores de duração absoluta e relativa do sistema vocálico

¹⁰ Maiores informações sobre o teste são encontradas em Purpura (2007).

do inglês durante o período de instrução do português. As células em amarelo correspondem às realizações mais longas.

Tabela 3. Médias (em milissegundos) e Desvios-Padrão das durações absolutas das vogais do inglês produzidas pelo participante do estudo longitudinal (n. de cada vogal = 18 por coleta, exceto [ɛ] com 15)

Coleta	[ɑ]	[æ]	[ʌ]	[ɛ]	[i]	[I]	[ɔ]	[u]	[ʊ]
1	152,4 DP:31,1	175,7 DP:31,9	152,8 DP:30,6	163,9 DP:33,6	115,8 DP:36,1	118,4 DP:34,1	169,06 DP:38,1	135,2 DP:25,4	115,4 DP:25,4
2	164,3 DP:25,7	195,2 DP:29,3	173,8 DP:38,8	184,2 DP:32,0	138,9 DP:37,95	137,3 DP:41,2	195,1 DP:45,1	150,9 DP:31,9	133,2 DP:32,7
3	170,9 DP:27,8	202,9 DP:30,3	169,7 DP:35,7	191,6 DP:39,5	147,6 DP:26,5	140,4 DP:41,05	216,2 DP:44,8	153,3 DP:28,9	134,8 DP:25,5
4	161,9 DP:31,0	187,6 DP:31,8	172,7 DP:25,2	182,8 DP:30,7	133,6 DP:25,4	139,3 DP:43,0	190,9 DP:33,3	151,9 DP:41,1	137,06 DP:20,5
5	180,6 DP:18,2	195,56 DP:22,45	171 DP:25,9	178,9 DP:29,0	138,8 DP:26,7	145,3 DP:32,5	195,9 DP:37,2	157,6 DP:26,7	146,8 DP:27,6

Tabela 4. Médias (em porcentagens) das durações relativas das vogais do inglês produzidas pelo participante do estudo longitudinal (n. de cada vogal = 18 por coleta, exceto [ɛ] com 15)

Monosílabos	[ɑ]	[æ]	[ʌ]	[ɛ]	[i]	[I]	[ɔ]	[u]	[ʊ]
Coleta 1	20,19 DP:2,6	22,73 DP:2,0	19,54 DP:1,4	19,99 DP:2,4	16,24 DP:1,4	17,68 DP:2,4	23,00 DP:2,3	17,46 DP:2,3	16,67 DP:1,2
Coleta 2	20,49 DP:2,95	24,14 DP:3,0	22,76 DP:1,6	22,87 DP:1,8	18,31 DP:2,6	20,42 DP:1,8	24,85 DP:2,6	20,16 DP:2,8	17,77 DP:2,7
Coleta 3	20,30 DP:2,1	24,65 DP:3,4	21,15 DP:4,1	23,37 DP:1,5	18,27 DP:1,1	19,73 DP:2,1	28,26 DP:2,7	19,40 DP:2,7	17,68 DP:1,7
Coleta 4	19,66 DP:3,9	22,63 DP:2,7	20,95 DP:1,85	21,57 DP:1,74	16,99 DP:1,9	20,07 DP:4,1	24,43 DP:2,4	19,07 DP:4,6	17,20 DP:2,7
Coleta 5	22,13 DP:1,9	23,17 DP:1,8	20,58 DP:1,8	21,03 DP:1,8	18,28 DP:3,0	18,71 DP:2,3	24,98 DP:2,7	18,60 DP:3,1	19,95 DP:2,6
Dissílabos	[ɑ]	[æ]	[ʌ]	[ɛ]	[i]	[I]	[ɔ]	[u]	[ʊ]
Coleta 1	12,71 DP:1,5	15,49 DP:1,25	12,91 DP:1,2	13,46 DP:1,7	8,64 DP:2,2	9,19 DP:1,4	13,38 DP:1,8	11,68 DP:1,4	10,14 DP:1,2
Coleta 2	14,98 DP:2,2	19,19 DP:2,45	14,38 DP:2,5	16,27 DP:1,5	11,55 DP:2,5	10,96 DP:1,7	16,80 DP:3,0	13,44 DP:1,1	12,89 DP:2,0
Coleta 3	15,71 DP:2,4	19,05 DP:1,6	15,32 DP:1,5	15,74 DP:1,3	13,28 DP:1,65	10,85 DP:1,6	18,15 DP:1,5	13,97 DP:1,7	12,47 DP:1,5
Coleta 4	15,18 DP:2,3	17,36 DP:1,4	16,26 DP:1,8	15,17 DP:1,2	12,09 DP:2,4	11,10 DP:2,6	16,47 DP:1,05	13,85 DP:0,96	14,09 DP:1,36
Coleta 5	16,27 DP:1,1	18,30 DP:1,3	15,67 DP:1,7	15,30 DP:1,3	12,22 DP:1,3	12,69 DP:2,5	16,33 DP:1,2	13,95 DP:1,2	13,26 DP:0,85

Os valores apresentados nas Tabelas 3 e 4 mostram uma tendência semelhante à L3 (português) em termos de duração da L2 (inglês): a terceira e a quinta coleta são as que dispõem de valores mais altos, o que representa uma maior duração vocálica, mas os valores encontram-se em flutuação e não apresentam uma gradiente ascensão. No caso do inglês, o participante realiza as vogais de forma mais alongada na segunda, terceira e quinta coletas.

Similarmente aos resultados referentes à duração absoluta encontrados no português do participante, a primeira coleta é a que dispõe dos valores mais baixos. Mesmo a segunda coleta, cujos valores, ainda que mais altos, são mais baixos do que os encontrados na terceira e na quinta coleta, apresenta valores mais altos do que as vogais do português deste aprendiz, de modo a indicar que o sistema vocálico do inglês dispõe de durações maiores do que o seu sistema de L3. Além disso, o período após a terceira coleta, que mostrou desestabilização do sistema em termos de eixos de F1 e de F2 (cf. PEREYRON, 2017), apresenta as vogais nas suas mais alongadas realizações, em ambas as línguas. Na L2, as vogais médias anteriores, como nas palavras 's[ɛ]t' e 'c[æ]t', e a média posterior, como na palavra 'c[ɔ]ffin', atingem suas maiores durações na terceira coleta, e o mesmo ocorre na L3: a média anterior aberta, como em 'qu[ɛ]pe', e a média posterior aberta, como em 'c[ɔ]po', apresentam as maiores durações neste momento do desenvolvimento do aprendiz. Já a vogal [ɑ] da L2 apresenta a maior duração na quinta coleta, seguida da terceira coleta, justamente nas duas etapas do desenvolvimento do aprendiz em que a referida vogal mostra maior aproximação, em termos de F1, da vogal [ɔ].

Dessa forma, a instrução referente às vogais abertas do português parece ter influenciado não somente as durações das vogais abertas do próprio português, mas também as do inglês do participante trilingue. As vogais abertas do sistema da L2 parecem ter sido desestabilizadas justamente no momento em que novas categorias foram formadas na L3 e, pela inter-relação entre os espaços acústicos do participante, as durações parecem ter sofrido efeitos de alongamento exatamente neste período de caos.

Os valores apresentados nas Tabelas 3 e 4 representam, mais uma vez, a não-linearidade, uma propriedade dos sistemas caóticos e dos sistemas complexos, uma vez que, ao observarmos os valores das coletas 1 a 5, percebemos que as durações absolutas não tendem a mostrar alteração crescentemente, mas aumentam da coleta 1 a 2 apenas. Depois da terceira coleta, as vogais mostram oscilações em termos de valores de duração, de modo a não apresentar comportamento linear. De acordo com Chan (2001), a complexidade emerge da inter-relação, interação e da interconectividade dos elementos do sistema e entre o sistema e seu ambiente. Tal fato implica que uma ação de uma das partes do sistema

influenciará todas as outras partes relacionadas, e não de uma maneira uniforme, o que representa o caráter não-linear desses sistemas. Chan (2001) também sugere que sistemas complexos exploram o seu espaço de possibilidades e criam novos padrões de relações. Tais sistemas funcionam melhor quando combinam ordem e caos em uma medida apropriada: estabilidade e instabilidade andam juntas e, conforme Larsen-Freeman (1997) afirma, a sistemacidade e a variabilidade não desafiam uma à outra, elas se alimentam.

Em síntese, a instrução formal com foco na articulação das vogais médias altas e médias baixas do português parece ter exercido influência tanto no sistema-alvo (L3), quanto no sistema de L2 do participante deste estudo, sobretudo com referência às vogais médias baixas anteriores (em termos de frequências e duração vocálica) e às vogais médias posteriores (em termos de duração vocálica). Há evidências para acreditar que, a partir da terceira coleta, que apresenta valores formânticos e de duração máximos, uma alteração em um sistema linguístico pôde alterar o outro, provavelmente devido a esse ponto culminante encontrado neste período do desenvolvimento, que, de tão alterado, parece afetar, também, a outra língua.

Tomaram-se por base teórica os autores adeptos à TSD e à Teoria da Complexidade, que argumentam a favor da interconexão entre os sistemas, bem como do fato de que as línguas do falante multilíngue pertencem a um mesmo sistema psicolinguístico. Tais aspectos levam, de fato, a influências múltiplas entre as línguas. Desse modo, a hipótese que regeu o estudo longitudinal até este ponto de análise foi confirmada em termos de alterações na L2 transferidas da L3. Cabe ainda investigar se essas alterações também exercem influências na L1 do participante, o espanhol.

A seguir, será apresentada a análise do sistema linguístico materno do participante deste estudo, a variedade mexicana de espanhol.

4.3 O espanhol

Quanto às durações das vogais do espanhol durante as cinco coletas, a duração absoluta das cinco vogais mostrou-se alterada, sendo essas realizadas mais brevemente nas duas primeiras coletas e mais alongadas nas duas últimas coletas, de modo a indicar que a instrução oferecida ao aluno (juntamente com outras variáveis, como o próprio contexto de imersão no português) possa ter exercido influência na duração vocálica do sistema materno. Tal ocorrido está em consonância com a premissa defendida neste trabalho, de que a transferência vocálica pode ocorrer da língua menos dominante em direção à língua mais dominante, isto é, a materna. Esses resultados corroboram os argumentos apontados por Grosjean (2013), que afirmam que a transferência é, no mínimo, bidirecional. A

Tabela 5 a seguir expõe os valores referentes à duração absoluta das vogais do espanhol do participante e a Tabela 6, os valores referentes à duração relativa. As células em amarelo correspondem às durações mais longas.

Tabela 5. Médias (em milissegundos) e Desvios-Padrão (DP) das durações absolutas das vogais do espanhol produzidas pelo participante do estudo longitudinal (n. de cada vogal= 18 por coleta, total de 90 por coleta)

	[a]	[e]	[i]	[o]	[u]
Coleta 1	110,55 DP:15,9	102,33 DP:16,0	74,16 DP:10,7	112,38 DP:18,7	93,16 DP:15,3
Coleta 2	110,83 DP:8,5	83,22 DP:9,1	68,83 DP:12,1	104,77 DP:9,9	80,55 DP:13,95
Coleta 3	123,5 DP:17,4	106,72 DP:18,1	84,44 DP:10,5	129,22 DP:13,2	103,27 DP:12,7
Coleta 4	115,38 DP:19,2	108,72 DP:18,7	91,5 DP:11,2	127,94 DP:14,2	105,66 DP:16,8
Coleta 5	142,55 DP:14,15	124,94 DP:14,5	97,88 DP:18,7	147,38 DP:19,45	118,88 DP:22,2

Tabela 6. Médias (em porcentagens) e Desvios-Padrão (DP) das durações relativas das vogais do espanhol produzidas pelo participante do estudo longitudinal (n. de cada vogal = 18 por coleta, total de 90 por coleta)

Dissílabos	[a]	[e]	[i]	[o]	[u]
Coleta 1	13,69 DP: 2,5	12,74 DP:2,5	8,83 DP:1,3	13,70 DP:2,1	12,37 DP:1,7
Coleta 2	14,45 DP:1,1	10,31 DP:1,1	9,04 DP:1,0	13,31 DP:1,3	11,14 DP:1,5
Coleta 3	15,99 DP:1,2	13,02 DP:2,3	9,63 DP:1,05	15,63 DP:1,3	12,72 DP:1,2
Coleta 4	15,65 DP:1,4	14,42 DP:2,0	11,76 DP:1,5	16,33 DP:1,3	13,66 DP:1,85
Coleta 5	16,97 DP:0,82	14,98 DP:1,7	11,84 DP:2,5	17,19 DP:2,3	15,04 DP:1,5
Trissílabos	[a]	[e]	[i]	[o]	[u]
Coleta 1	11,25 DP:1,5	10,66 DP:1,6	7,81 DP:1,3	11,89 DP:2,4	8,79 DP:2,0
Coleta 2	11,91 DP:0,85	9,65 DP:0,9	7,64 DP:1,7	11,34 DP:1,3	7,88 DP:1,5
Coleta 3	12,02 DP:1,5	10,98 DP:1,8	9,50 DP:1,35	13,27 DP:1,6	10,21 DP:1,5
Coleta 4	11,56 DP:1,9	10,49 DP:1,65	9,36 DP:1,4	13,52 DP:1,3	10,58 DP:1,6
Coleta 5	13,94 DP:1,45	12,39 DP:1,4	9,90 DP:1,5	15,15 DP:2,1	10,80 DP:2,3

As durações apresentadas nas Tabelas 5 e 6 anteriores, no que diz respeito a primeira e a segunda coleta, sofrem uma diminuição nos valores, com exceção da vogal baixa [a], que, na segunda coleta, apresenta uma média de duração muito próxima à da primeira coleta. As durações das demais vogais diminuem consideravelmente entre a primeira e a segunda coleta, possivelmente devido a um período de instabilidade do sistema, já que a segunda coleta ocorreu após apenas 4 sessões de instrução do português. Tal fato pode ser interpretado como se o sistema vocálico da L1, com a intensificação da L3, estivesse entrando em um estado de caos, característico dos sistemas dinâmicos adaptativos complexos.

No restante das coletas, as durações absolutas das vogais do espanhol sofrem um aumento gradual, o que provavelmente tenha ocorrido pela exposição ao português bem como à instrução acelerada, da mesma forma que ocorreu com as durações dos sistemas maternos dos participantes residentes no Brasil do estudo transversal. Esse ocorrido está em consistência com a premissa dinâmica de que qualquer alteração pode alterar o restante do(s) sistema(s), e ainda com a afirmação de Bybee (2010), de que o contexto e a frequência de uso das palavras são fortes condicionadores para o detalhe fonético. A autora, ademais, sugere que a repetição do input a que o aprendiz é exposto, ou seja, a frequência com que as palavras individuais ou sequências são empregadas, afeta a representação mental, e, nesse sentido, o uso da língua inclui, além do processamento, os usos e as interações sociais.

Cabe ainda ressaltar que, ao observarmos os valores da quinta coleta na Tabela 5 anterior, as durações vocálicas do participante na L1 são idênticas, ou pelo menos muito próximas, às durações da primeira coleta da L3 (português), conforme apresentado na Tabela 1 anterior: o aprendiz apresenta uma média de 142,9 ms para [a]; 125,7 ms para [e]; 97,2 ms para [i]; 135,5 para [o]; e 117,2 ms para [u]. Essas semelhanças mostram que a duração mais longa do participante na L1 (quinta coleta) corresponde à duração mais curta de que ele dispõe na L3 (primeira coleta, período pré-instrução). Esse achado indica que, mesmo antes da instrução, o informante já produzia, na L3, uma vogal mais longa do que aquela da sua L1 (possivelmente em função da exposição à fala de Porto Alegre), e essas vogais ficaram ainda mais longas com a instrução. Nesse sentido, essas “alterações” na duração das vogais da L1 do participante apresentam um valor intermediário, que não correspondem à duração do português, tampouco à do espanhol. Esses valores correspondem às durações que o participante tinha na L3 antes da instrução. Assim, a instrução parece conduzir o aprendiz a aumentar seus valores de duração na L3 e, também, na L1.

Porém, na L1, esse aumento se dá até um ponto definido, que corresponde à duração que tal aprendiz tinha na L3, antes da instrução (que, por sua vez, já era uma categoria híbrida). Além disso, os valores mais longos na L1, mesmo antes da instrução da L3, são provavelmente resultado da exposição do aprendiz ao português, língua do país em que ele vive, e essa exposição contínua já havia causado alguma modificação na língua-materna antes mesmo do período de instrução.

Finalmente, o período de maior instabilidade (que corresponde à terceira coleta, em que se podem observar diversas alterações nos sistemas do informante) parece ter exercido, mesmo que mais amenamente, efeitos sobre a duração das vogais do sistema materno, pelo menos em relação à [a] e à [o]. Conforme já discutido, as vogais do sistema materno apresentaram as maiores durações na quinta coleta, momento em que o participante já se encontrava no Brasil há um semestre e havia participado de todas as sessões da instrução formal. No entanto, foi na terceira coleta que as vogais [a] e [o] apresentaram as durações mais altas, depois da quinta coleta. Devido à interconexão entre os sistemas, a desestabilização do sistema de L3 do aprendiz parece ter sido transferida para todos os outros aspectos por nós investigados.

5 Conclusão

Com base nos resultados descritos, no que concerne ao desenvolvimento da L2 e da L3 do participante deste estudo, há uma considerável inter-relação entre os sistemas, além da interconexão que a L2 e a L3 mostraram com a L1. Há bastante variabilidade nos valores, o que é esperado, dentro do arcabouço teórico utilizado nesta pesquisa, devido, entre diversos fatores, à não-linearidade dos sistemas complexos. A variabilidade possivelmente deriva do desenvolvimento acelerado causado pela instrução intensiva, o que sugere que o processo de desenvolvimento linguístico constitui, de fato, um sistema dinâmico adaptativo complexo. Os valores de duração da terceira coleta na língua de instrução mostraram-se bastante distintos dos demais, o que sugere o caráter não-linear no processo de desenvolvimento vocálico, uma propriedade da TSD e da Teoria da Complexidade. Os valores encontrados na terceira coleta da L3, ademais, parecem refletir um momento de desestabilização do sistema que se reflete, também, na L2 do aprendiz.

A L1 do participante também se mostrou alterada, de modo que é possível sugerir que a transferência entre os sistemas linguísticos também pode ocorrer da língua mais recente à língua mais entrincheirada. Considerando uma perspectiva dinâmica e complexa, esta pesquisa fundamenta-se na pressuposição de que é necessário rejeitar a noção unidi-

recional de transferência linguística. A partir de tais resultados, pôde-se afirmar que uma L2 deve ser considerada como um sistema híbrido, que carrega características tanto da L1 quanto da L3. A L3, sob essa perspectiva, carrega aspectos das L1 e da L2. Além disso, a própria L1, ainda que de forma mais moderada, sofre influência dos outros sistemas que se estabelecem neste espaço fonológico comum. É rejeitada, portanto, a noção de direção singular quanto à transferência das L1 e L2 na L3, de modo que se possa assumir, assim, uma transferência multidirecional.

Referências

ALVES, M. A. *Estudo dos parâmetros acústicos relacionados à produção das plosivas do Português Brasileiro na fala adulta: análise acústico-quantitativa*. Tese (Doutorado em Letras). Florianópolis: Universidade Federal de Santa Catarina, 2015.

BARBOZA, C. L. F. *Efeitos da Palatalização das Alveolares do Português Brasileiro no Percurso de Construção do Inglês como Língua Estrangeira*. Tese (Doutorado em Letras). Fortaleza: Universidade Federal do Ceará, 2013.

BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M.; CROFT, W.; ELLIS, N.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T. *Language is a Complex Adaptive System: Position Paper*. *Language Learning*, 59 (1), 2009 p. 1-26.

BLANK, C. A. *A influência grafo-fônico-fonológica na produção oral e no processamento de priming em multilíngues: uma perspectiva dinâmica*. Tese (Doutorado em Letras). Pelotas: Universidade Católica de Pelotas, 2013.

_____. A transferência grafo-fônico-fonológica em multilíngues falantes de português, espanhol, e inglês: uma abordagem dinâmica. In: ALVES, U. K. (Org.). *Aquisição Fonético-Fonológica de Língua Estrangeira: Investigações Rio-Grandenses e Argentinas em Discussão*. Campinas, SP: Pontes Editores, 2016, p. 209-228.

BOERSMA, P.; WEENINK, D. *Praat*, Versão 5.4.08. Disponível em: <http://www.praat.org>. Acesso em: 9 nov. 2017.

BRITO, K. S. *Influências interlinguísticas na mente multilíngue: perspectivas psicolinguísticas e (psico)tipológicas*. Tese (Doutorado em Letras). Curitiba: Universidade Federal do Paraná, 2011.

BYBEE, J. *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.

CELCE-MURCIA, M.; BRINTON, D. M.; GOODWIN, J. M. *Teaching pronunciation: a reference for teachers of English to speakers of other languages*. Nova Iorque: Cambridge University Press, 1996.

CENOZ, J. The Effect of Linguistic Distance, L2 Status and Age on Cross-linguistic Influence in Third Language Acquisition. In: CENOZ, J.; HUFEISEN, B.; JESSNER, U. (Ed.). *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Bristol: Multilingual Matters, 2001, p. 8-20.

CHAN, S. *Complex Adaptive Systems*. Disponível em: <http://web.mit.edu/esd.83/www/notebook/Complex%20Adaptive%20Systems.pdf>. Acesso em: 21 fev. 2019.

DE BOT, K. Introduction: Second Language Development and a Dynamic Process. *The Modern Language Journal*, v. 92, n. ii, p. 166-178, 2008.

_____; LOWIE, W.; VERSPOOR, M. A Dynamic System Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, v. 10, n. 1, p. 7-21, 2007.

_____; LARSEN-FREEMAN, D. Researching second language development from a dynamic systems perspective. In: VERSPOOR, M. H.; DE BOT, K.; LOWIE, W. (Ed.). *A Dynamic Approach to Second Language Development: Methods and Techniques*. Amsterdam: John Benjamins Publishing Company, 2011, p. 5-24.

_____; LOWIE, W.; THORNE, S. L.; VERSPOOR, M. Dynamic System Theory as a comprehensive theory of second language development. In: P. García Mayo, J. Gutierrez Mangado, & M. Martínez Adrián (Ed.). *Contemporary perspectives on second language acquisition*. John Benjamins Publishers, 2013, p. 167-189.

ELLIS, N. C. The Emergence of Language as a Complex Adaptive System. In: SIMPSON, J. (Ed.). *Routledge Handbook of Applied Linguistics*. Londres: Routledge / Taylor Francis, 2011, p. 654-667.

FEIDEN, J. A.; PEROZZO, R. V.; FINGER, I.; FONTES, A.B.A.L. Percepção de vogais médias altas e médias baixas do português brasileiro por falantes de espanhol rioplatense em tarefa de discriminação categórica. In: ALVES, U. K. (Org.). *Aquisição Fonético-Fonológica de Língua Estrangeira: Investigações Rio-Grandenses e Argentinas em Discussão*. Campinas, SP: Pontes Editores, 2016, p. 85-103.

FRANCESCHINI, R. Sprachadoption: der Einfluss von Minderheitensprachen auf die Mehrheit, oder: Welche Kompetenzen der Minderheitensprachen haben Minderheitensprachen? *Bulletin Suisse de Linguistique Appliquée*, v. 69, n. 2, p. 137-153, 1999.

GROSJEAN, F. Bilingualism: a Short Introduction. In: GROSJEAN, F.; LI, P. *The psycholinguistics of Bilingualism*. Londres: Wiley-Blackwell, 2013, p. 13-31.

HERDINA, P.; JESSNER, U. *A Dynamical Model of Multilingualism: Perspectives of Change in Psycholinguistics*. Trowbridge: Cromwell Press Ltd., 2002.

JESSNER, U. Teaching third languages: Findings, trends and challenges. *Language teaching*, v. 41, n. 1, p. 15-56, 2008.

LARSEN-FREEMAN, D. Chaos/complexity science and second language acquisition. *Applied linguistics*, v. 18, n. 2, p. 141-165, 1997.

_____. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied linguistics*, v. 27, n. 4, p. 590-619, 2006.

_____. Complex, dynamic systems: A new transdisciplinary theme for applied linguistics? *Language Teaching*, v. 45, n. 2, p. 202-214, 2012.

_____. Complexity theory: a new way to think. *Revista Brasileira de linguística aplicada*, v. 13, n. 2, p. 369-373, 2013. Disponível em: <http://dx.doi.org/10.1590/S1984-63982013000200002>. Acesso em: 5 jun. 2016.

_____; CAMERON, L. *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press, 2008.

LIMA JR., R. M. A necessidade de dados individuais e longitudinais para análise do desenvolvimento fonológico de l2 como sistema complexo. *ReVEL*, v. 14, n. 27, p. 203-225, 2016a.

LIMAJR, Ronaldo M. Análise longitudinal de vogais do inglês-L2 de brasileiros. *Gradus-Revista Brasileira de Fonologia de Laboratório*, v. 1, n. 1, p. 145-176, 2016b.

LINDBLOM, B. Phonetic universals in vowel systems. In: OHALA, J.; JAEGER, J. (Ed.). *Experimental Phonology*. Florida: Academic Press, 1986, p. 13-44.

PEREYRON, L. *A produção vocálica por falantes de Espanhol (L1), Inglês (L2) e Português (L3): uma perspectiva dinâmica na (multi)direcionalidade da transferência linguística*. Tese (Doutorado em Letras). Porto Alegre: Universidade Federal do Rio Grande do Sul, 2017.

PURPURA, J. *Conceptualizing and measuring meaning in SLA Research, Second Language Research Forum (SLRF)*, University of Illinois, Champaign-Urbana. 2007. Disponível em <http://oxford.britanica-edu.org>. Acesso em: 6 ago. 2016.

SANTOS, G. *Percepção e produção das vogais médias do espanhol por falantes do português brasileiro*. Tese (Doutorado em Letras). Pelotas: UCPEL, 2014.

_____; RAUBER, A. S. *Percepção e produção das vogais médias do espanhol/LE*. In: ALVES, U. K. (Org.). *Aquisição Fonético-Fonológica de Língua Estrangeira: Investigações Rio-Grandenses e Argentinas em Discussão*. Campinas: Pontes, 2016, p. 151-178.

SILVA, A. H. P. *A variável tempo nos estudos de aquisição*. In: BRAWERMAN-ALBINI, A.; CASTRO GOMES, M. L. *O jeitinho brasileiro de falar inglês*. Pontes, 2014a. p. 33-47.

SILVA, A. H. P. Organização temporal de encontros vocálicos no Português Brasileiro e a relação entre Fonética e Fonologia. *Letras de Hoje*, v. 49, n. 1, p. 11-18, 2014b.

VAN GELDER, T.; PORT, R. It's about time: an overview of the dynamics approach to Cognition. In: PORT, R.; VAN GELDER, T. (Ed.). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge: The MIT Press, 1995, p. 1- 43.



Data de submissão: 30/01/2018

Data de aceite: 18/11/2018

EFEITOS DE INSTRUÇÃO EXPLÍCITA NA AQUISIÇÃO DA ENTOAÇÃO EM ESPANHOL/L2

EFFECTS OF EXPLICIT INSTRUCTION
IN THE ACQUISITION OF INTONATION IN SPANISH/L2

Rhanya Rafaella Rodrigues | [Lattes](#) | rhanyarafaella@gmail.com
Instituto Federal Goiano - Campus Ceres | CAPES-MEC

Elena Ortiz Preuss | [Lattes](#) | ortizpreuss.e@gmail.com
Universidade Federal de Goiás

Resumo: A entoação é um elemento complexo e importante para a comunicação, entretanto, trata-se de um aspecto que tende a ser ignorado ou abordado de forma implícita, no contexto de aquisição de L2 (ASL) (AKERBERG, 2011; MORENO, 2002; DIAS, 2015). Em vista disso, e considerando o baixo número de investigações sobre a aquisição de elementos suprasegmentais em segunda língua (L2), esta pesquisa tem como objetivo analisar os efeitos da instrução explícita na aquisição de enunciados declarativos e interrogativos totais em espanhol/L2. Além disso, espera-se verificar se os estudantes apresentam maior desempenho em enunciados declarativos ou interrogativos totais. Para isso, realizou-se um estudo semiexperimental, com aprendizes brasileiros de língua espanhola, em nível intermediário-avançado. A pesquisa contou com 19 participantes que foram divididos em quatro grupos, com base em dois fatores: i) condição (experimental ou controle); e ii) contexto (graduação ou curso livre de idiomas). Para a geração de dados foram usados os seguintes instrumentos: i) questionário, que possibilitou o delineamento do perfil dos participantes; e ii) pré- e pós-testes linguísticos, que abrangeram tarefas de compreensão e produção (controlada, semi-controlada e livre) dos enunciados-alvo. As análises estatísticas não apontaram efeitos significativos da instrução explícita, embora a intervenção pareça ter auxiliado na minimização de efeitos de similaridade, no caso dos declarativos, e na produção dos interrogativos totais. Notou-se também que os enunciados interrogativos totais parecem ser mais difíceis que os declarativos, para os aprendizes brasileiros.

Palavras-chave: Instrução explícita; Entoação; Espanhol/L2.

Abstract: Intonation is a complex and important element for communication. Nevertheless, this aspect tends to be overlooked or approached in an implicit way, in the context of second language acquisition (SLA). Having this in mind and considering the low number of researches on acquisition of suprasegmental features in second language (L2), this study aims at analyzing effects of explicit instruction on acquisition of declarative and total interrogative utterances in Spanish/L2. Furthermore, we expect to verify if students perform higher in total interrogative and declarative utterances. To that end, we conducted a semi-experimental study, with Brazilian learners of Spanish, at an intermediate-advanced level. The 19 participants were divided into four groups, based on two factors: i) condition (experimental or control); and ii) context (graduation course or independent language course). For data generation, we used the following instruments: i) questionnaire, which enabled the design of the profile of students; and ii) pre and post linguistic tests, which encompassed tasks of comprehension and production (controlled, semi-controlled and free) of target-utterances. Conducted statistical analyses did not indicated significant effects of explicit instruction, although the intervention seemed to help in the minimization of similarity effects, in cases of declarative utterances, and in the production of total interrogative ones. In addition, total interrogative utterances appear to be harder than declarative ones for Brazilian learners.

Keywords: Explicit instruction; Intonation; Spanish/L2.

Introdução

A entoação tem um papel fundamental na comunicação, independentemente do idioma em que se esteja comunicando, posto que é responsável por expressar tanto o tipo de enunciado (afirmação, negação, interrogação), quanto a intenção e os sentimentos do falante (esclarecer uma dúvida, expressar surpresa, demonstrar medo, etc.). Além disso, trata-se de um dos primeiros aspectos fonético-fonológicos que é adquirido pelo ser humano, o que interfere em sua aquisição em uma segunda língua (L2), pois há maior tendência à transferência dos padrões entoacionais da primeira língua para a L2 (AKERBERG, 2011; DIAS, 2015; PINTO, 2009; SEMINO, 2014).

Na aquisição de entoação em espanhol/L2 por aprendizes brasileiros, destacam-se, segundo as autoras, dois tipos de enunciados como os que apresentam maior distinção entre o espanhol e o português brasileiro (PB), a saber: os declarativos e os interrogativos totais. A diferença contida nesses dois enunciados pode causar problemas comunicativos

entre os interlocutores, pois a transferência dos padrões melódicos da L1 para a L2 influencia diretamente no tipo de enunciado que está sendo produzido (por exemplo, uma interrogação pode ser entendida como uma afirmação).

Alves (2012) argumenta que há um procedimento metodológico denominado instrução explícita que pode auxiliar o aprendiz a perceber as principais diferenças entre os elementos segmentais e suprasegmentais da L1 e da L2. Conforme o autor, a instrução explícita de aspectos fonético-fonológicos está em consonância com uma proposta de ensino comunicativo da pronúncia (CELCE-MURCIA et al., 2010), que se desenvolve em cinco etapas diferentes: i) discriminação auditiva; ii) instrução explícita; iii) prática controlada; iv) prática semi-controlada; v) prática livre.

Note-se que a aquisição da entoação é uma área com baixo número de pesquisas, e que apresenta muitas questões em aberto, sobretudo no que se refere à aquisição de espanhol por aprendizes brasileiros (DIAS, 2015; MORENO, 2002; SANZ, 2005; ORTIZ-PREUSS; SANZ, 2016). Considerando-se essa lacuna no campo da ASL, esta pesquisa tem como objetivo geral observar os efeitos da instrução explícita na aquisição de enunciados declarativos e interrogativos totais em espanhol/L2 e analisar o desempenho dos participantes nos testes de percepção e produção dos enunciados-alvo.

Para isso, desenvolveu-se uma pesquisa semiexperimental, de cunho quantitativo. Na geração de dados, foram usados os seguintes instrumentos: i) Questionário de Histórico e Proficiência Linguística (QHPL) (MARIAN; BLUMENFELD; KAUSHANSKAYA, 2007) e ii) Pré-Teste e Pós-Teste Linguístico. No total, o estudo teve 19 participantes, divididos em quatro grupos, com base no seu contexto de estudo (graduação ou curso livre de espanhol) e no recebimento ou não instrução explícita (experimental e controle).

Este artigo está organizado em quatro seções. Inicialmente, expõe-se o referencial teórico que norteou este estudo. Posteriormente, descreve-se o método utilizado para a geração de dados; consecutivamente, apresentam-se os resultados e discussões. Finalmente, constam algumas considerações finais.

Referencial Teórico

A Instrução Explícita em L2

Durante muitos anos, os estudos na área de aquisição de segunda língua (ASL) enfocaram o resultado do processo de aquisição, ou seja, o produto. Entretanto, nas últimas décadas, a forma como a L2 é aprendida passou a ser o foco das pesquisas (FONTANA;

LIMA, 2009). Sanz (2005) aponta que a partir dessa nova perspectiva, três hipóteses surgiram no campo da ASL, são elas: i) a hipótese do *input* compreensível (KRASHEN, 1985); ii) a hipótese do *output* compreensível (SWAIN, 1995); e iii) a hipótese interacionista (LONG, 1991).

Dentre essas hipóteses¹, a proposta de Long, na década de 90, reconfigurou os procedimentos pedagógicos da época, pois o autor apresenta uma visão mais branda sobre o ensino comunicativo, que integra os aspectos comunicativos da língua ao ensino de estruturas linguísticas. Segundo Lyrio (2012), três propostas metodológicas são conceituadas, baseadas no foco dado à forma e ao significado, a saber: i) foco nas formas, que enfoca o ensino de elementos estruturais da língua, sem estabelecer relação entre forma-função; ii) foco no significado, que ignora o ensino de aspectos linguísticos, seu foco é o desenvolvimento da competência comunicativa; e iii) foco na forma (FonF), que visa ensinar tanto aspectos estruturais quanto abranger elementos da comunicação, ou seja, são procedimentos pedagógicos que chamam a atenção dos aprendizes para determinados aspectos da língua, de uma forma comunicativa e significativa para o aprendiz (ORTIZ-PREUSS, 2005; LYRIO, 2012; VIDAL, 2007).

No entanto, discute-se que o foco na forma concebido por Long (1991) limita-se a uma abordagem reativa, ou seja, é pontual e só acontece quando o professor nota alguma dificuldade do aprendiz. A instrução focada na forma (IFF)², por outro lado, permite a realização de abordagens proativas, nas quais a intervenção pedagógica é realizada antes que os alunos apresentem dificuldades (ELLIS, 2005).

Há diferentes possibilidades de procedimentos metodológicos a partir da IFF. Segundo Ellis (2005), pode-se realizar práticas de instrução explícita (dedutiva e indutiva) e implícita, produções mais ou menos controladas e livres, e *feedback* explícito ou implícito. Em vista disso, nota-se que a distinção explícito e implícito é importante nesse contexto.

De acordo com Alves (2004), Aquino (2012) e Ortiz-Preuss (2005), o conhecimento explícito é aprendido, controlado e verbalizável, o que não significa que se refira a um conhecimento metalinguístico. Ademais, esse conhecimento permite o acesso à consciência. O conhecimento implícito, por sua vez, é mais intuitivo e adquirido de forma incidental, manifesta-se através do uso, de modo automático e não-verbalizável, trata-se de

¹ Devido ao escopo deste artigo, não serão detalhadas as outras hipóteses. Para o leitor interessado, sugere-se a leitura de Sanz (2005).

² Salienta-se que, nesta pesquisa, optou-se pelo uso da instrução focada na forma, uma vez que é mais abrangente e por se adequar à proposta de intervenção usada para a geração de dados.

um conhecimento não consciente (ALVES, 2004; AQUINO, 2012; ORTIZ-PREUSS, 2005).

No campo da ASL, a (não) conversão desses conhecimentos é bastante investigada. Discute-se três possibilidades de interface entre esses dois tipos de conhecimento. A primeira é a Hipótese da Não-Interface que sustenta que um conhecimento não pode ser convertido em outro, visto que se localizam em áreas diferentes no cérebro (ALVES, 2012; AQUINO, 2012; ORTIZ-PREUSS, 2005; DEKEYSER, 2009). Por outro lado, a segunda opção, conhecida como a Hipótese da Interface Forte afirma que a prática exaustiva da estrutura-alvo promove a conversão de um conhecimento em outro. Por fim, a Hipótese da Interface Fraca defende que o conhecimento pode se tornar conhecimento implícito, através da instrução explícita. Argumenta-se que são necessárias diferentes oportunidades de acesso ao *input* e de práticas comunicativas e contextualizadas, além disso, deve-se considerar a maturidade cognitiva do aprendiz (AQUINO, 2012; ORTIZ-PREUSS, 2005).

Dentre as diferentes possibilidades de intervenção pedagógica, a instrução explícita é considerada benéfica para a ASL (ALVES, 2004; ALVES; ZIMMER, 2005; ORTIZ-PREUSS, 2005; ELLIS, 2005). Segundo os autores, esse tipo de instrução objetiva direcionar a atenção do aprendiz, de forma explícita, para diferentes aspectos do *input*.

A instrução explícita pode facilitar a formação do conhecimento explícito e favorecer o automonitoramento, feito pelo aluno (ALVES, 2012). O autor esclarece que não se trata de um procedimento que ensina aspectos estruturais da L2, de modo descontextualizado, mas que garante o acesso a esses elementos de forma comunicativa e contextualizada.

Salienta-se que a instrução explícita auxilia o aprendiz a perceber elementos que poderiam não ser notados, durante o processo de ASL. Estudos têm apontado vantagens, com efeitos duradouros, para grupos que recebem instrução explícita, quando comparados a estudantes que não têm acesso a esse tipo de intervenção (ALVES, 2004; ORTIZ-PREUSS, 2005). Contudo, é importante destacar que também há evidências negativas para os efeitos da instrução explícita (cf. ELLIS, 2005).

A instrução explícita, conforme exposto, pode ser usada com diferentes aspectos linguísticos e pragmáticos. Nesta pesquisa, opta-se pelo uso da instrução explícita para ensino da entoação em espanhol/L2. Em vista disso, na próxima subseção, será apresentada uma breve discussão sobre a entoação em espanhol.

Aquisição de Entoação em Espanhol/L2

A entoação é um aspecto suprasegmental que envolve aspectos linguísticos, paralinguísticos e extralinguísticos (nível fonológico) e elementos como duração, intensidade e tom (nível fonético) (RAE, 2011; SOSA, 1999). Segundo Semino (2014), espanhol e português brasileiro são línguas entoativas, pois a entoação marca sentidos linguísticos e pragmáticos.

De acordo com Sosa (1999), a entoação abrange quatro premissas, a saber: i) é significativa; ii) é sistemática; iii) é característica; e iv) possui uma unidade melódica. Além disso, a entoação parece cumprir, ao menos, três funções: a primeira é conhecida como *expressiva*, refere-se à possibilidade de transmissão de sentimentos e emoções; a segunda função é a *focalizadora* que possibilita ao falante enfatizar alguma informação da mensagem; e a função *demarcativa* que se relaciona à segmentação e organização dos enunciados (AKERBERG, 2011; DIAS, 2015; PRIETO, 2003; RAE, 2011; SEMINO, 2014; SOSA, 1999).

Como foi exposto por Sosa (1999, p. 30, tradução nossa), um dos aspectos da entoação é apresentar elementos específicos de padrão entoacional, conforme as diferentes variedades.

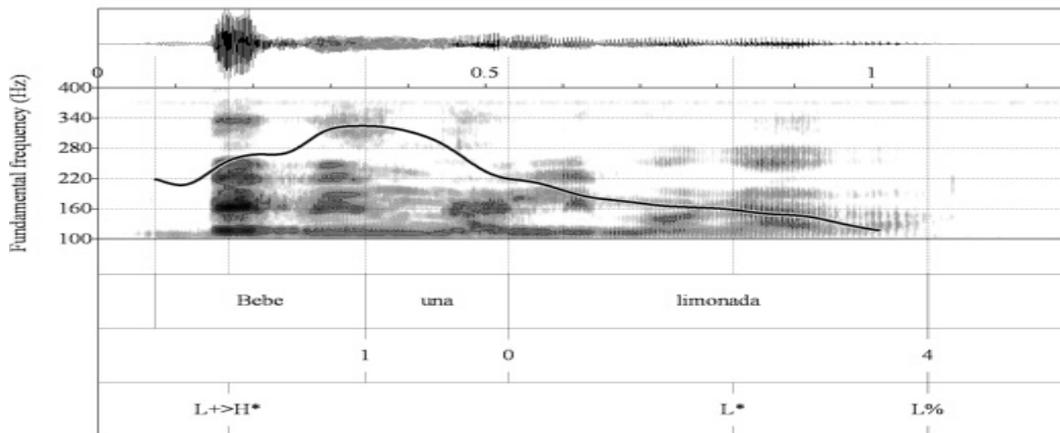
No caso específico do espanhol, nem todos os dialetos usam os mesmos padrões entoacionais para os mesmos efeitos, por isso, pode-se afirmar que ao mesmo tempo em que há uma quantidade de características em comum a todos os dialetos do espanhol, existem também fenômenos de entoação privativos a alguns dialetos que não se produzem em outros.

Salienta-se que, embora se esteja de acordo com o teórico no que diz respeito às variedades, neste estudo, optou-se pelos padrões entoacionais propostos pela RAE (2011)³. Com base na RAE (2011, p. 455, tradução nossa), os enunciados declarativos são usados “[...] quando se explica um pensamento ou se dá uma informação ao seu interlocutor”. Em espanhol, esses enunciados são produzidos com um tom médio inicial (L*+H), após a primeira sílaba tônica há subidas melódicas (L*H%) até o tonema descendente (↓) (L*L%). Sosa (1999) compara o movimento descendente ao da bola de tênis, que apresenta uma diminuição progressiva ao tocar o chão. Em PB, o movimento inicial é ascendente (↑), seguido por descenso contínuo do tom (conhecido como linha de declinação) e finalizado por um tonema descendente (↓) (PINTO, 2009).

³ A opção por esse padrão se justifica pela necessidade de operacionalizar a instrução explícita, tendo em vista que se contava com um tempo reduzido para a intervenção. Destaca-se que essa escolha deu-se, também, pelo fato de apresentarem os padrões com elementos mais recorrentes nos dois tipos de enunciados.

Note-se que os padrões desses enunciados são similares em espanhol e PB, visto que apresentam tonema descendente (\downarrow), o aspecto que diferencia os dois enunciados é o descenso progressivo, em espanhol, conforme Figura 1, e a linha de declinação, em PB (Figura 2).

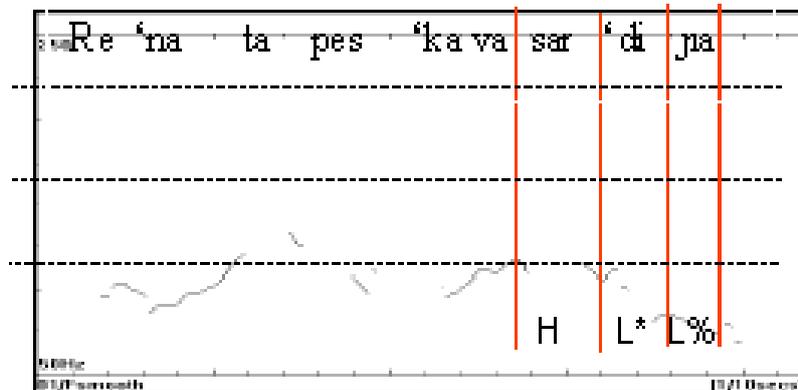
Figura 1. Padrão dos enunciados declarativos em Espanhol



Fonte: Pietro e Roseano (2009).

Figura 2. Padrão enunciados declarativos PB

Renata pescava sardinha.



Fonte: Pinto (2009, p. 42).

Os enunciados interrogativos podem ser caracterizados como parciais⁴, que são introduzidos por partículas interrogativas e requerem complemento de informações, ou interrogativos totais, que demandam respostas do tipo sim/não (RAE, 2011; SEMINO,

⁴ Os enunciados interrogativos parciais apresentam padrão entoacional similar entre as duas línguas (SEMINO, 2014).

2014; SOSA, 1999). Neste estudo, optou-se pelos interrogativos totais, porque apresentam maior distinção no padrão entoacional entre espanhol e PB.

A maior diferença no padrão melódico desse enunciado entre as línguas parece estar no tonema e pós-tonema, pois o espanhol possui contorno final ascendente (↑), enquanto o PB tem uma curva circunflexa (^), ou seja, há um tom baixo, seguido de um ascenso e finalizado por outro tom baixo (AKERBERG, 2011; DIAS, 2015; PINTO, 2009). Nas duas línguas há um tom alto inicial, seguido por um descenso, porém, conforme discutido, o contorno final é divergente entre as línguas. Dias (2015) e Pinto (2009) destacam que, no que concerne ao espanhol, o tom inicial tende a ser produzido com uma elevação tonal, que é concebida por Sosa (1999) como predisposição inicializadora. As autoras acrescentam que esse tom tende a ser mais alto em espanhol do que em PB.

Akerberg (2011) e Moreno (2002) problematizam que embora a entoação seja um fenômeno complexo e extremamente importante para a comunicação, em sala de aula, tende a ser ignorado ou abordado de forma muito implícita, predominando o ensino de aspectos segmentais (pronúncia). Dias (2015) aponta que embora haja uma quantidade considerável de estudos sobre os enunciados aqui enfocados, as pesquisas se concentram em aspectos relativos à transferência prosódica e a análises descritivas/contrastivas. Em vista disso, esta pesquisa visou a analisar efeitos da instrução explícita na aquisição da entoação em espanhol/L2. Na próxima seção será exposto o método utilizado para o desenvolvimento deste estudo.

Método

Esta pesquisa, de natureza quantitativa semiexperimental, justifica-se pelo baixo número de estudos sobre a aquisição de aspectos suprasegmentais em L2 (AKERBERG, 2011; DIAS, 2015), e objetiva observar os efeitos de instrução explícita na aquisição de enunciados declarativos e interrogativos totais em L2; e analisar se há diferenças entre os testes de compreensão e produção dos enunciados-alvo.

Dois contextos diferentes foram testados: um curso de licenciatura em Letras – Espanhol, em que a instrução explícita foi aplicada no sétimo período (correspondente ao penúltimo)⁵ que enfocava exclusivamente a habilidade oral; e um curso livre de idiomas, no qual a instrução explícita foi aplicada a estudantes do sexto nível do curso de espanhol (equivalente ao último semestre do curso).⁶

⁵ Salienta-se que a disciplina tinha acabado de ser reformulada, de modo que os alunos egressos não tiveram instruções explícitas específicas sobre a entoação. Por isso, foram convidados a integrar o grupo controle egressos do curso de Licenciatura em Letras – Espanhol.

⁶ Como nesse contexto não havia previsão de desenvolvimento do conhecimento metalinguístico sobre o espanhol, participaram do grupo controle alunos que não tinham recebido instrução explícita sobre o tema.

O estudo contou com 19 participantes⁷, todos estudantes brasileiros de espanhol, que possuíam nível intermediário-avançado e que concordaram em participar voluntariamente⁸ da pesquisa. Os participantes foram divididos em quatro grupos, a saber: i) Grupo Experimental Graduação (GEG); ii) Grupo Controle Graduação (GCG); iii) Grupo Experimental Curso Livre (GEC); e iv) Grupo Controle Curso Livre (GCC). O quadro exposto a seguir apresenta as principais características de cada grupo.

Quadro 1. Descrição dos Participantes da Pesquisa

SIGLA	TIPO	CONTEXTO	Nº DE PARTICIPANTES	SEXO	MÉDIA DE IDADE
GEG	Experimental	Graduação	5	5 F	27,5 anos
GCG	Controle	Graduação	4	4 F	23,25 anos
GEC	Experimental	Curso Livre	5	4 F 1 M	22,4 anos
GCC	Controle	Curso Livre	5	4 F 1 M	34,8 anos

Para a geração de dados foram utilizados dois instrumentos diferentes. Primeiramente, os participantes responderam o QHPL (MARIAN; BLUMENFELD; KAUSHANSKAYA, 2007), que é um questionário usado para o delineamento do perfil do participante (histórico linguístico, frequência de uso/contato linguístico e proficiência linguística). O questionário é composto por 23 perguntas sobre histórico de aquisição e proficiência linguística (auto-avaliada).

O pré-teste e pós-teste⁹ foram aplicados, respectivamente, antes e depois do recebimento da instrução explícita, com vistas a verificar se a instrução alterou, de algum modo, o desempenho dos participantes dos grupos experimentais. Os testes eram compostos por quatro tarefas distintas:

- i) compreensão, na qual os participantes deveriam ouvir os enunciados e indicar se eram declarativos (15 enunciados), interrogativos totais (15 enunciados) ou negativos¹⁰ (10 enunciados);

⁷ Tratou-se de uma amostra de conveniência, o que justifica o baixo número de participantes.

⁸ Pesquisa aprovada pelo CEP, sob nº CAAE: 49155915.50000.5083.

⁹ Ressalta-se que os testes apresentaram a mesma estrutura, alterando-se apenas as frases, áudios e imagens entre um teste e outro. Para a aplicação do teste, foi utilizado o programa E-prime 2.0, que registrou tempo de resposta e acurácia nos enunciados de compreensão. Nos testes de produção, também foi utilizado um gravador de voz para registro das respostas dos participantes.

¹⁰ Os enunciados negativos foram usados como distratores, a fim de verificar se os participantes estavam ou não atentos ao experimento.

- ii) produção controlada, em que os participantes deveriam ler em voz alta os enunciados que apareciam na tela do computador (10 declarativos, 10 interrogativos totais e 5 negativos);
- iii) produção semi-controlada, em que os participantes deveriam produzir 5 enunciados declarativos e 5 enunciados interrogativos totais, considerando a imagem e a pista que aparecia na tela (+ declarativos ou ? interrogativos totais); e
- iv) prática livre, na qual os participantes produziam enunciados, com base nas figuras que apareciam na tela, sendo que, nesse caso, o tipo de enunciado era escolhido pelos integrantes da pesquisa.

O pré-teste e o pós-teste foram intercalados por um período de instrução explícita. No total, a instrução explícita foi realizada em quatro aulas (5,6 horas/aula), em ambos os grupos. Todas as atividades se relacionavam ao tema “viagem”, apontado como relevante pelos alunos. Inicialmente, foi solicitado aos estudantes que reagissem às frases lidas pela professora, que alternava aleatoriamente a entoação usada (declarativa ou interrogativa total) e observava a reação dos alunos. Em seguida, foi feita a instrução explícita relativa aos enunciados-alvo, para isso foram apresentados espectrogramas e materiais didáticos sobre o tema; os participantes também tiveram acesso a uma síntese¹¹ do quadro elaborado por Semino (2014) que apresenta as principais distinções entre os enunciados em espanhol e PB.

Após o recebimento da instrução explícita, os participantes realizaram atividades de discriminação auditiva¹². Na atividade aplicada aos dois grupos os participantes enumeravam os enunciados, conforme a ordem em que o ouviam. A segunda tarefa consistia em ouvir enunciados iguais e assinalar qual era o seu tipo (declarativo ou interrogativo total).

Na sequência, deu-se início às atividades de prática. A primeira atividade foi de prática controlada em que os alunos recebiam dois enunciados (1 declarativo e 1 interrogativo total) que deveriam ser lidos com base na entoação indicada. Após a leitura das frases, os discentes as gravaram no programa *PRAAT*¹³, o que os permitiu visualizar a cur-

¹¹ Salienta-se que os alunos do contexto de graduação também tiveram acesso a uma explicação metalinguística sobre o conteúdo, tendo em vista o contexto em que estão inseridos. Para isso, utilizou-se o capítulo sobre entoação (SEMINO, 2014).

¹² Destaca-se que os participantes do GEC realizaram uma atividade a mais de instrução explícita, que se justifica pela menor exposição diária ao *input* auditivo na L2, visto que os alunos do GEG tinham a maioria das aulas ministradas em língua espanhola.

¹³ Elaborado por Paul Boersma e David Weenink. Disponível em: <<http://www.fon.hum.uva.nl/praat/>>

va melódica produzida e compará-la com o que era esperado para cada tipo de enunciado.

Na prática semicontrolada, os alunos realizaram a atividade “*Busca a alguien que...*”, nesse caso, os participantes receberam uma lista com 10 frases relacionadas a viagens e deveriam encontrar, entre os colegas, pessoas que atendessem a cada um dos itens. Desse modo, os participantes alternavam entre perguntas e afirmações.

A prática livre, para os participantes do GEG, consistiu na leitura de textos sobre “Mochilões” e sobre a prática de hospedagem em albergues e *hostels* e, posteriormente, na realização de um debate sobre o tema. Já os participantes do GEC entrevistaram um salvadorenho que vivia em Goiânia havia 3 anos. Ressalta-se que embora as atividades de prática livres tenham sido diferenciadas, pressupõe-se que não houve impacto no efeito da instrução, posto que a estrutura “perguntas e respostas” foi mantida.

A tabulação e análise de dados foram baseadas nas características dos dados. Os dados referentes ao QHPL foram tabulados no programa Excel e analisados no SPSS 1.8¹⁴. Os dados relativos ao teste de compreensão auditiva foram tabulados no programa *E-prime 2.0*, avaliando-se a acuidade (ACC)¹⁵ e o tempo de reação (TR). Após a tabulação, os dados foram analisados estatisticamente no programa SPSS 1.8.

Para a análise de dados de produção, inicialmente, foi realizada a fragmentação dos áudios e minimização de ruídos através do programa *Audacity 2.1*. Posteriormente, os dados foram analisados no *PRAAT*, a fim de que se observasse a curva melódica produzida pelos participantes. Salienta-se que foi utilizado o método métrico-autossegmental para atribuição de padrões melódicos (SOSA, 1999). Isto posto, em caso de enunciados produzidos em conformidade com o padrão, atribuiu-se o escore (1), que indicava acerto. Por sua vez, enunciados que não correspondiam ao modelo entoacional abordado eram considerados errados, utilizando-se o escore (0), para fins de análise estatística.

Na seção seguinte, serão expostos os principais dados e discussões acerca dos resultados.

Resultado e Discussões

As análises de dados foram realizadas no programa SPSS 1.8. Inicialmente, fez-se uma análise descritiva e exploratória, para que fosse verificada a normalidade da amostra. Os testes de Kolmogorov-Smirnov e Shapiro-Wilk indicaram que os dados não esta-

¹⁴ *Statistical Package for the Social Sciences*.

¹⁵ Para fins de avaliação, atribui-se o valor 0 para acertos e 1 para erros.

vam em uma distribuição normal. Deste modo, aplicou-se os testes não-paramétricos de Wilcoxon, para a comparação entre pré-teste e pós-teste (dois momentos intrasujeitos), e o teste de Mann-Whitney (dois momentos intersujeitos), que permitiu a comparação de médias entre grupos (controle e experimental) e contextos (graduação e curso livre).

Em um primeiro momento, serão apresentadas as análises referentes aos testes de compreensão, seguidas das análises dos testes de produção.

Testes de Compreensão

A apresentação dos resultados seguirá a seguinte sequência: i) testes de compreensão; e ii) testes de produção controlada, semi-controlada e livre. Em um primeiro momento, serão expostos os dados dos participantes do contexto de graduação; seguidos daqueles do contexto de curso livre e, por fim, serão comparados os resultados dos dois contextos.

Os testes de compreensão continham 40 questões (15 declarativas, 15 interrogativas totais e 10 livres) e os participantes tinham que indicar o tipo dos enunciados (A- declarativo, N- negativo e ?- interrogativo). Na sequência, serão expostos os dados de média (M) e tempo de reação (TR), bem como o desvio padrão (DP).

Tabela 1. Resultados dos testes de compreensão do contexto de graduação

	GEG (N=5)		GCG (N=4)	
	PRÉ M (DP)	PÓS M (DP)	PRÉ M (DP)	PÓS M (DP)
Afirmativas_acc (N=15)	14,20 (0,83)	14,60 (0,89)	13,75 (0,50)	15,00 (0,00)
Interrogativas_acc (N=15)	13,80 (1,78)	14,20 (1,78)	14,25 (0,50)	15,00 (0,00)
Afirmativas_TR	3138,85 (274,22)	2853,28 (299,02)	3284,18 (325,90)	2853,88 (400,48)
Interrogativas_TR	2864,74 (254,95)	2761,47 (303,74)	2800,76 (355,62)	2540,68 (208,51)

Nota: GEG= Grupo Experimental Graduação; GCG= Grupo Controle Graduação; acc= acurácia; TR= tempo de reação em milissegundos; M= média; DP= desvio padrão.

Note-se na Tabela 1 que, no pré-teste, o GCG foi o grupo que obteve a menor e a maior média de acurácia na compreensão das afirmativas (M= 13,75, no pré-teste e M= 15,00, no pós-teste), o que demonstra que o GCG apresentou maior variabilidade entre

os testes. O GEG não alterou muito a média de acurácia entre os testes (14,20, no pré-teste e 14,60, no pós-teste), e os índices de DP (0,83, no pré-teste e 0,89, no pós-teste).

Também se percebe que ambos os grupos aumentaram as suas médias de acertos no pós-teste (GEG= 14,60 e GCG= 15,00) e mantiveram o índice de DP, que foi do GEG, o que parece indicar que a instrução pode ter efeitos diferenciados entre os participantes. Salienta-se que, como o padrão enunciativo dos declarativos não apresenta muita diferença entre a L1 e a L2, é possível que o aprendiz tenha se embasado em sua L1.

Na compreensão das interrogativas, a menor média foi apresentada pelo GEG (M=13,80) e a maior média foi do GCG (M=15,00). Ademais, novamente, o GEG apresentou maior índice de DP (1,78), indicando maior heterogeneidade na média de acertos entre os participantes. As maiores médias de ambos os grupos ocorreram no pós-teste (GEG= 14,20 e GCG= 15,00), destacando-se que não houve diferença significativa nas médias que indicasse efeitos da instrução no GEG, porém, a variabilidade dos DPs pode ser um indício de que a instrução tenha afetado os alunos de modos diferentes.

Os dados foram submetidos a análises estatísticas que não mostraram diferenças significativas entre grupos no contexto de graduação. Em outras palavras, nesse caso, a instrução explícita pode ter auxiliado no desempenho dos aprendizes, mas não foi um fator responsável pelo desempenho superior do GEG em relação ao GCG.

O tempo que os participantes levaram para executar as respostas também foi avaliado. Na realização da compreensão de enunciados afirmativos, o GEG teve a menor média no pós-teste (M=2853,28), enquanto que o GCG apresentou a maior média (M=3284,18), ou seja, os participantes do GEG foram mais rápidos para responder. O GEG também se diferenciou do GCG, no que se refere aos índices de DP, sendo que o GEG apresentou menor variabilidade entre participantes nos testes se comparados ao GCG (GEG=274,22, no pré-teste e 299,02, no pós-teste; GCG=325,90, no pré-teste e 400,48, no pós-teste).

Por outro lado, na compreensão dos interrogativos, o GCG obteve a menor média (M=2540,68, no pós-teste), e o GEG foi o grupo com a maior média (M= 2864,74). Entretanto, deve-se salientar que o GEG apresentou diminuição do TR no pós-teste (M=2761,47). Quanto ao DP, o GEG aumentou os índices no pós-teste (de 254,94 para 303,74), e o GCG os diminuiu (de 355,62 para 208,51). As análises estatísticas, assim como nos testes de acurácia, não apontaram diferença significativa entre os grupos. Ou seja, a instrução explícita não levou os participantes a serem mais rápidos que os participantes do GCG.

Ainda que o GCG tenha obtido médias de acuidade maiores que as do GEG não houve diferença estatisticamente significativa entre os grupos. A hipótese inicial era de que os participantes, de ambos os grupos, obteriam maiores médias de acurácia na compreensão de afirmativos, considerando-se a proximidade da curva melódica desses enunciados nas duas línguas. Porém, o GCG apresentou médias de acurácia maiores na compreensão de interrogativos totais no pré-teste, além de obter a mesma média para os dois tipos de enunciados no pós-teste. A partir desses resultados, conjectura-se que a compreensão de interrogativos totais não seja tão difícil para os participantes desse grupo.

Esperava-se que a acurácia do GEG aumentasse no pós-teste, em decorrência da instrução explícita. Os resultados não corroboraram essa hipótese, posto que ambos os grupos aumentaram seus escores, não havendo diferenças significativas entre os mesmos. Deste modo, pressupõe-se que o aumento da acurácia esteja atrelado à familiaridade com a tarefa.

Assim como nos testes de acurácia, ambos os grupos diminuíram suas médias de TR no pós-teste, em outras palavras, todos os participantes foram mais rápidos para responder o pós-teste do que o pré-teste. Como não houve diferenças significativas entre os grupos, supõe-se que, nesse caso, os resultados também tenham sido causados pela familiaridade com a tarefa.

Frisa-se que, ao contrário do que era esperado, os participantes foram mais rápidos para responder os enunciados interrogativos totais do que os declarativos, nos dois testes. Discute-se que esses resultados possam ter sido influenciados pelo fato de que os enunciados interrogativos totais apresentam o tom inicial mais alto do que os declarativos (DIAS, 2015; PINTO, 2009; SOSA, 1999). Dessa forma, os participantes podem ter percebido que se tratava de um enunciado interrogativo total no início ou na última sílaba tônica, que sofre um aumento tonal considerável. Nos declarativos, os participantes, provavelmente, escutaram todo o enunciado para confirmar seu tipo, visto que as diferenças são mais sutis, nesse caso.

Note-se que os dois grupos apresentaram aumento de acurácia e diminuição de TR no pós-teste, em ambos os enunciados. Sendo assim, a instrução explícita parece não ter sido a principal variável que tenha influenciado nas médias obtidas, de modo que outros aspectos possam interferir na compreensão dos enunciados aqui estudados.

A Tabela 2, exposta a seguir, apresenta os resultados nos testes de compreensão dos participantes do contexto de curso livre.

Tabela 2. Resultados dos testes de compreensão do contexto de curso livre

	GEC (N=5)		GCC (N=5)	
	PRÉ M (DP)	PÓS M (DP)	PRÉ M (DP)	PÓS M (DP)
Afirmativas_acc (N=15)	14,20 (0,83)	14,80 (0,44)	14,00 (1,00)	14,80 (0,44)
Interrogativas_acc (N=15)	13,40 (0,89)	13,40 (1,81)	14,20 (0,83)	14,20 (1,30)
Afirmativas_TR	3109,79 (288,07)	2715,44 (279,35)	3070,53 (153,91)	2807,91 (94,76)
Interrogativas_TR	2697,64 (223,11)	2576,29 (213,05)	2781,75 (197,54)	2553,23 (55,32)

Verifica-se, na Tabela 2, que na compreensão dos afirmativos, o grupo que obteve a menor média, no geral, foi o GCC, no pré-teste (M=14,00), enquanto que os dois grupos obtiveram médias idênticas no pós-teste (M=14,80). Com base nas médias de acurácia e nos índices de DP do pós-teste, ambos os grupos acertaram mais e tiveram um comportamento mais homogêneo. O aumento de médias nos grupos experimental e controle indica que a instrução explícita parece não ter sido fundamental, posto que os participantes que não a receberam também aumentaram a acuidade no pós-teste.

Nos enunciados interrogativos, os grupos não alteraram os escores de acurácia do pré-teste para o pós-teste (GEC= 13,40 e GCC=14,20). Conjectura-se que a instrução explícita não tenha sido suficiente para que os participantes do GEC se sobressaíssem aos do GCC, considerando que este último grupo teve média maior que o grupo experimental. Entretanto, os dois grupos aumentaram o DP do pré-teste (GEC= 0,89 e GCC= 0,83) para o pós-teste (GEC= 1,81 e GCC= 1,30). Ou seja, embora as médias tenham se mantido, o desempenho foi mais heterogêneo, principalmente no GEC, o que pode sugerir que a instrução tenha tido efeitos diferentes entre esses participantes.

Na compreensão de enunciados declarativos e interrogativos totais, a instrução explícita parece não ter contribuído para um maior índice de acertos, pois, conforme os resultados, as médias dos grupos foram idênticas nos enunciados declarativos, e o GCC obteve média maior que a do GEC nos enunciados interrogativos totais. Além disso, destaca-se que não houve diferenças estatisticamente significativas entre os grupos.

Nas análises do TR, percebe-se, conforme a Tabela 2, que na compreensão de declarativos o GEC foi o grupo que mais alterou a média entre os testes, sendo o grupo mais lento no pré-teste (M=3109,79), e o mais rápido no pós-teste (M= 2715,44). No que se refere ao DP, o GCC foi o grupo com desempenho mais homogêneo (DP= 94,76, no pós-

-teste). Cabe ressaltar que o GCC também foi mais rápido em suas respostas no pós-teste ($M=2807,91$), se comparado ao pré-teste ($M=3070,53$).

Nos enunciados interrogativos totais, o GCC foi o grupo com maior variação entre os testes, obtendo a maior média no pré-teste ($M=2781,75$), e a menor no pós-teste ($M=2553,23$). Este grupo também diminuiu consideravelmente os índices de DP do pré-teste ($197,54$) para o pós-teste ($55,32$), o que indica que os participantes, além de mais rápidos, foram mais homogêneos no pós-teste.

Os testes de Wilcoxon indicaram diferenças significativas no TR do GEC entre o pré-teste e o pós-teste nos enunciados declarativos ($Z=-2,023, p<0,04$), e nos interrogativos totais ($Z=-2,023, p<0,04$), sugerindo que a instrução explícita tenha levado os participantes a serem mais rápidos no pós-teste.

No geral, nota-se que nos enunciados declarativos os dois grupos obtiveram maiores escores de acurácia no pós-teste, sendo que as médias foram iguais para os dois grupos. Na compreensão de interrogativos totais, os grupos mantiveram as médias nos dois testes, sendo que o GCC apresentou o maior número de acertos. Entretanto, não houve diferenças estatisticamente significativas entre os grupos.

A hipótese de que os participantes apresentariam maior acurácia nos enunciados declarativos foi parcialmente confirmada, já que, exceto o GCC no pré-teste dos declarativos em comparação com os interrogativos totais, as médias dos declarativos foram maiores que as dos interrogativos, o que parece indicar que estes últimos sejam mais difíceis para os estudantes desse contexto.

Assim como no contexto de graduação, supunha-se que os participantes do grupo experimental apresentariam desempenho mais acurado que os do grupo controle, no pós-teste. Esta hipótese não foi corroborada, dado que os grupos tiveram médias iguais na compreensão de declarativos, e os escores do GCC foram maiores que os do GEC nos enunciados interrogativos totais. Esses dados demonstram que, aparentemente, a instrução explícita não surtiu efeitos na compreensão dos enunciados-alvo, embora os índices de DP no GEC pareçam indicar que a instrução possa ter afetado aos participantes de forma diferenciada.

Em relação ao TR, os dois grupos foram mais rápidos no pós-teste, nos dois tipos de enunciado. Nesse caso, a instrução explícita parece estar associada à diminuição do TR, com base nas análises estatísticas.

Novamente, a hipótese de que os participantes seriam mais rápidos na compreensão de declarativos do que de interrogativos totais não se confirmou, pois, em geral, os participantes dos dois grupos foram mais rápidos nos interrogativos totais. Presume-se

que a rapidez na compreensão dos interrogativos totais seja atribuída à tendência inicializadora, que se refere ao aumento tonal no início desse tipo de enunciado (SOSA, 1999).

Apesar de a instrução explícita não ter contribuído diretamente na acurácia dos enunciados, os participantes do GEC foram mais rápidos na compreensão tanto dos declarativos, quanto dos interrogativos totais. Tais resultados podem ser atribuídos à instrução explícita, posto que foram explicadas diferenças no padrão entoativo dos dois tipos de enunciados.

Na subseção seguinte, serão expostos os dados relativos aos testes de produção tanto no contexto de graduação, quanto no de curso livre.

Testes de Produção

Além dos testes de compreensão, também se analisou a produção dos enunciados-alvo. Os testes continham três tarefas diferentes, a saber: i) produção controlada (10 declarativos, 10 interrogativos totais e 5 negativas); ii) produção semi-controlada (5 declarativos e 5 interrogativos totais); e iii) produção livre (6 enunciados). Na tabela a seguir são expostas as médias de acurácia e DP nos testes de produção do contexto de graduação.

Tabela 3. Resultados da acurácia nos testes de produção (Graduação)

	GEG (N=5)		GCG (N=4)	
	PRÉ M (DP)	PÓS M (DP)	PRÉ M (DP)	PÓS M (DP)
AfC (N=10)	8,20 (1,30)	8,20 (1,30)	9,25 (0,50)	8,75 (1,50)
IntC (N=10)	7,60 (1,81)	7,80 (1,64)	7,50 (3,10)	7,80 (2,06)
AfS (N=5)	2,75 (1,50)	2,80 (1,09)	3,00 (1,15)	3,75 (1,25)
IntS (N=5)	2,20 (0,83)	2,25 (1,50)	1,33 (1,15)	2,75 (1,25)
PL (N=6)	3,20 (1,09)	4,00 (1,82)	3,75 (1,25)	4,75 (1,25)

Na produção de afirmativos controlados, o GEG manteve a média em ambos os testes (M= 8,20) e o GCG diminuiu a média do pré-teste (M=9,25) para o pós-teste (M=8,75), além de aumentar o índice de DP (1,50). No que se refere aos enunciados interrogativos totais, o GCG foi o grupo que apresentou maior variação entre os testes (M= 7,50, no pré-teste, e M= 8,25, no pós-teste), além de ter diminuído o DP (2,06). Destaca-se que nesse tipo de enunciado, o GEG também apresentou aumento na média do pós-teste (M= 8,25).

Os resultados indicam que os dois grupos aumentaram a acurácia no pós-teste, no caso dos interrogativos totais, desse modo, percebe-se que a instrução explícita parece não justificar a média de acertos do GEG. Por outro lado, nos enunciados declarativos, houve uma queda nos escores do GCG, e o GEG manteve a sua média, por essa razão, supõe-se que a instrução explícita possa ter contribuído para que isso acontecesse. Esses dados indicam que aparentemente houve transferência negativa do padrão entoativo das línguas, o que pode ser justificado pela tendência a que a aquisição seja afetada, sobretudo nos casos de produção, quando se trata de línguas próximas (ORTIZ-PREUSS, 2014).

Na produção de afirmativos semi-controlados, o GEG apresentou um pequeno aumento das médias do pré-teste ($M= 2,75$) para o pós-teste ($M= 2,80$). O GCG foi o grupo que apresentou o maior aumento na quantidade de acertos entre os dois testes ($M= 3,00$, no pré-teste e $M= 3,75$, no pós-teste). Porém, cabe ressaltar que enquanto o GEG diminuiu o índice de DP (Pré= 1,50 e Pós= 1,09), o GCG teve um leve aumento no DP (1,15, no pré-teste e 1,25, no pós-teste). Em relação aos enunciados interrogativos semicontrolados, os dois grupos aumentaram as médias no pós-teste ($M= 2,25$, no GEG, e $M= 2,75$, no GCG). O maior aumento na média de acertos do pré-teste para o pós-teste foi do GCG ($M=1,42$). O GEG, por sua vez, aumentou consideravelmente o índice de DP entre os testes (0,83, no pré-teste e 1,50, no pós-teste), o que pode ser indício de que a instrução tenha surtido efeitos diferentes entre os participantes desse grupo. No geral, percebe-se que os participantes obtiveram maiores médias de acurácia na produção de enunciados declarativos do que na de interrogativos totais, corroborando a hipótese de que este último tipo de enunciado é mais difícil aos brasileiros (AKERBERG, 2011; DIAS, 2015).

Conforme a Tabela 3, no que se refere à produção livre, nota-se que ambos os grupos aumentaram a média de acertos no pós-teste ($M= 3,75$, no GEG, e $M= 4,75$, no GCG), sendo que a maior média de aumento foi apresentada pelo GCG ($M=1,00$). Em relação à variabilidade apresentada pelos participantes, o GEG aumentou o DP no pós-teste (1,82), já o GCG manteve o índice de DP nos dois testes (1,25). Esses dados podem sugerir, novamente, que a instrução tenha afetado aos participantes de modo diferente.

Destaca-se que embora o GCG tenha apresentado, em geral, médias maiores que as do GEG nos três tipos de tarefas, as análises estatísticas não indicaram diferenças significativas entre os grupos.

No que se refere à produção controlada de enunciados, os dois grupos aumentaram as médias nos interrogativos totais. Por outro lado, nos declarativos, o GEG manteve a média nos dois testes, e o GCG apresentou diminuição na média de acertos no pós-teste.

Julga-se que o aumento de médias, no caso dos interrogativos, tenha sido provocado pela familiaridade com a tarefa. Sobre os declarativos, supõe-se que a instrução explícita possa ter neutralizado os efeitos que a similaridade no padrão entoativo possa ter causado. Apesar de os participantes não terem aumentado os escores nos declarativos, observa-se que a hipótese de que a acurácia seria maior nesse tipo de enunciado foi confirmada, visto que, de modo geral, as médias foram maiores nos enunciados declarativos do que nos interrogativos totais.

Na tarefa de produção semi-controlada, os participantes aumentaram a média de acertos nos dois tipos de enunciados, ainda que os escores tenham sido maiores nos declarativos. Mais uma vez, os resultados favorecem a hipótese de que os interrogativos totais têm um padrão entoativo mais difícil de ser produzido do que os declarativos. Frisasse que, embora o GEG tenha aumentado a acurácia no pós-teste, a instrução explícita parece não ter sido fundamental, uma vez que o GCG também acertou mais no pós-teste.

Os resultados da produção livre se assemelham aos outros dois tipos de condição, dado que o aumento da acurácia do GEG, no pós-teste, não pode ser associado à instrução explícita, pois o GCG também apresentou médias superiores, nesse teste. Nota-se que a hipótese de que a instrução explícita levaria aos participantes do grupo experimental a acertarem mais que o grupo controle não foi corroborada, já que o GCG se sobressaiu ao GEG.

A Tabela 4, exposta a seguir, apresenta os resultados dos participantes do contexto de curso livre, nos testes de produção.

Tabela 4. Resultados da acurácia nos testes de produção (Curso Livre)

	GEC (N=5)		GCC (N=5)	
	PRÉ M (DP)	PÓS M (DP)	PRÉ M (DP)	PÓS M (DP)
AfC (N=10)	8,20 (1,48)	8,20 (1,09)	8,40 (1,34)	7,60 (2,07)
IntC (N=10)	8,20 (1,30)	9,60 (0,54)	6,60 (2,40)	8,00 (2,34)
AfS (N=5)	3,60 (1,51)	3,40 (1,51)	3,60 (1,94)	3,40 (1,67)
IntS (N=5)	3,20 (0,44)	3,80 (1,30)	2,20 (1,64)	2,20 (0,44)
PL (N=6)	3,40 (1,67)	5,20 (0,83)	5,40 (0,89)	3,40 (0,54)

Em relação à produção controlada de afirmativos, o GCC diminuiu a média de acertos do pré-teste ($M=8,40$) para o pós-teste ($M=7,60$), além disso, esse grupo aumentou o DP ($1,34$, no pré-teste e $2,07$, no pós-teste). Ou seja, os participantes erraram mais no pós-teste, mas variaram mais entre si. O GEC, por sua vez, preservou os escores nos dois testes ($M= 8,20$) e diminuiu o índice de DP ($1,09$, no pós-teste). Verifica-se que, nessa condição, a instrução explícita não surtiu efeitos. Nos interrogativos totais, os dois grupos aumentaram a acurácia no pós-teste ($M= 9,60$, no GEC e $M= 8,00$, no GCC), considerando que os dois grupos tiveram aumento na média de acertos no pós-teste, não é possível indicar que os resultados do GEC tenham sido causados, indispensavelmente, pela instrução explícita. Sobre o DP, observa-se que o GEC diminuiu a variabilidade no desempenho entre os participantes ($1,30$, no pré-teste e $0,54$, no pós-teste), o que parece evidenciar que a instrução tenha provocado efeitos mais similares no GEC.

Com base nesses resultados, presume-se que o aumento de médias na produção de interrogativos possa estar relacionado à familiaridade com a tarefa, enquanto que nos declarativos a instrução possa ter colaborado na manutenção das médias dos declarativos, no GEC. Conjectura-se que, assim como no contexto de graduação, a similaridade entre línguas tenha causado efeitos negativos na produção desses enunciados (ORTIZ-PREUSS, 2014).

Na produção de afirmativos semi-controlados, as médias dos dois grupos se mantiveram tanto no pré-teste ($M= 3,60$, em ambos), quanto no pós-teste ($M= 3,40$, em ambos). Constata-se que os participantes tiveram uma pequena diminuição na acurácia, no pós-teste. Em relação ao DP, o GCC diminuiu seus índices ($DP= 1,97$, no pré-teste para $DP= 1,67$, no pós-teste), e o GEC não apresentou variação no DP entre os testes ($DP=1,51$). Esses resultados indicam que a instrução não surtiu efeitos, pois o GEC diminuiu a média de acertos e manteve o comportamento entre os participantes. No que tange aos interrogativos semi-controlados, o GEC aumentou os escores no pós-teste ($M=3,80$), e o GCC manteve a mesma média nos dois testes ($M=2,20$). Com base no aumento do DP no GEC ($DP= 0,44$, no pré-teste e $DP= 1,30$, no pós-teste), conjectura-se que a instrução tenha provocado efeitos diferentes nos participantes desse grupo.

Na produção de enunciados semicontrolados é possível perceber que os dois grupos diminuíram a média de acertos nos declarativos, e que apenas o GEC aumentou os escores na produção de interrogativos totais.

Em relação aos enunciados livres, nota-se que o GCC apresentou uma diminuição considerável nas médias do pré-teste ($M = 5,40$) para o pós-teste ($M= 3,40$). Por outro

lado, o GEC apresentou aumento na média de acertos ($M = 3,40$, no pré-teste e $M = 5,20$, no pós-teste), além de apresentar diminuição na variação do desempenho dos participantes ($DP = 0,83$, no pós-teste). Salienta-se que, conforme o Teste de Wilcoxon, a diminuição no GCC foi estatisticamente significativa ($Z = -2,041$, $p < 0,04$).

Na produção de enunciados controlados, constata-se que os dois grupos aumentaram os escores nos interrogativos totais. Nos declarativos, por outro lado, o GEC manteve a mesma média nos dois testes, já o GCC diminuiu os escores no pós-teste. Supõe-se que a familiaridade com a tarefa tenha influenciado na produção dos interrogativos, ao passo que, nos declarativos, a instrução parece ter contribuído para que o GEC minimizasse o efeito da similaridade. Nessa tarefa, a hipótese de que os participantes obteriam maiores escores na produção de declarativos não se confirmou, tendo em vista que os dois grupos acertaram mais na produção de interrogativos totais.

Sobre os enunciados semicontrolados, nota-se que, no caso dos declarativos, os dois grupos diminuíram os escores no pós-teste. Nos interrogativos totais, por seu turno, o GEC aumentou a acurácia, e o GCC manteve a mesma média nos dois testes. Novamente, a hipótese de que os participantes teriam mais facilidade com os enunciados declarativos não foi corroborada, haja vista que o GEC apresentou médias mais altas nos interrogativos totais. Em compensação, o GCC confirmou a hipótese, dado que as médias desse grupo foram maiores nos enunciados declarativos do que nos interrogativos totais.

Na produção livre, observa-se que o GEC teve um aumento considerável na acurácia, no pós-teste, e o GCC, por sua vez, diminuiu significativamente os seus escores. Esses resultados sinalizam que a instrução explícita possa ter influenciado na produção do GEC, uma vez que a diminuição no GCC foi estatisticamente significativa. Contudo, ressalta-se que as análises não indicaram diferenças significativas para os efeitos da instrução explícita.

Com base na maior média de acertos do GEC, no geral, a hipótese de que o grupo experimental apresentaria maior desempenho que o grupo controle foi parcialmente confirmada, nesse contexto. Entretanto, frisa-se que o aumento nos escores do GEC não pode ser atribuído à instrução explícita, uma vez que não houve diferenças estatisticamente significativas, e o GCC também apresentou aumento, em praticamente todas as tarefas, no pós-teste.

Na seção seguinte, serão expostas as principais considerações sobre os resultados encontrados nesta pesquisa.

Considerações finais

A realização deste estudo justificou-se pela carência de pesquisas que investiguem a aquisição de aspectos suprasegmentais em L2. Com isso, esta pesquisa visou a contribuir com o escopo sobre entoação em espanhol/L2, principalmente no que se refere a enunciados declarativos e interrogativas totais.

Esperava-se observar os efeitos da instrução explícita na aquisição desses dois tipos de enunciados. Os resultados indicaram que, no geral, os participantes dos grupos experimentais (GEG e GEC) acertaram mais no pós-teste, em ambos os testes (compreensão e produção). Contudo, no contexto de graduação, os escores dos participantes do grupo controle foram superiores aos dos participantes do grupo experimental. Destaca-se que as análises estatísticas não apontaram diferenças significativas para o efeito da instrução explícita. De modo que ainda que o GEG tenha aumentado a acurácia, no pós-teste, a instrução não foi suficiente para que esse grupo obtivesse médias significativamente maiores que as do GCG.

No contexto de curso livre, os participantes do GEC aumentaram as médias no pós-teste, em quase todos os testes (compreensão e produção). Além disso, na maioria das tarefas, o GEC obteve desempenho mais acurado que o GCC. Entretanto, assim como no contexto de graduação, o aumento dos escores no pós-teste não pode ser atribuído à instrução explícita, visto que não houve diferenças estatisticamente significativas.

Esperava-se também verificar o desempenho dos participantes nos testes de compreensão e produção, considerando-se os dois tipos de enunciados analisados. Em relação à compreensão, verifica-se que, em ambos os contextos, os participantes tiveram desempenho mais acurado e mais lento, nos enunciados declarativos. Por outro lado, nos enunciados interrogativos totais, os grupos erraram mais, mas responderam mais rapidamente. Esses resultados indicam que a compreensão de declarativos é mais fácil, mas o seu processamento é mais lento. Os interrogativos totais, por sua vez, são mais difíceis, porém são processados de forma mais rápida. Essa agilidade no processamento de interrogativos totais parece ser atribuída à tendência inicializadora, em que o tom médio e o tonema final são produzidos com tom mais alto que os dos declarativos (SOSA, 1999).

No que concerne à produção, os grupos tendem a apresentar maior dificuldade nos interrogativos totais do que nos declarativos. No entanto, o GEC obteve médias maiores na produção de interrogativos totais, mantendo as médias nos afirmativos controlados, e diminuindo os escores no semicontrolados. Infere-se que a instrução explícita possa ter contribuído para que os participantes se mantivessem mais atentos

aos enunciados interrogativos totais, por apresentarem estrutura mais divergente da sua L1.

Contudo, presume-se que outros fatores, tais como as diferenças individuais de cada aprendiz, também possam estar relacionadas ao fato de a instrução explícita não ter surtido efeitos estatisticamente significativos. Sanz (2014, 2005) argumenta que os procedimentos pedagógicos devem ser estudados em interação com as variáveis individuais dos aprendizes, de modo que se consiga explicar a grande variabilidade em seus níveis de proficiência.

Cabe destacar que o estudo apresentou algumas limitações, como o tamanho da amostra, um curto período para a aplicação da instrução, a impossibilidade de aplicação de um pós-teste tardio, entre outras. Sugere-se, a partir do exposto, que estudos com maior duração e amostras mais populosas sejam realizados. Destaca-se que a realização de pesquisas que abranjam a aquisição de aspectos fonético-fonológicos da L2 é fundamental, principalmente no caso do espanhol, que ainda conta com um escopo bastante reduzido. Além disso, é importante que os procedimentos pedagógicos sejam analisados em associação com as diferenças individuais, a fim de que se possam explicar lacunas encontradas nos testes linguísticos.

Referências

AKERBERG, M. Aquisição da pronúncia: a entoação em afirmações e perguntas sim/não. *Trabalhos em Linguística Aplicada*, Campinas, v. 50, n.1, p. 153-167, 2011.

ALVES, U. K.; ZIMMER, M. C. Perceber, notar e aprender: uma visão conexional da consciência do aprendiz na aquisição fonológica da L2. *Revista Virtual de Estudos da Linguagem – ReVEL*, Ano 3, n. 5, 2005.

_____. A explicitação dos aspectos fonético-fonológicos da L2: teoria e pesquisa na sala de aula. In: LAMPRECHT, R.R. (Org.). *Consciência dos sons da língua: subsídios teóricos e práticos para alfabetizadores, fonoaudiólogos e professores de língua inglesa*. 2 ed. Porto Alegre: EDIPUCRS, 2012. p. 211-230.

_____. O papel da instrução explícita na aquisição fonológica do inglês como L2: evidências fornecidas pela Teoria da Otimidade. 2004. 335f. Dissertação (Mestrado em Letras) - Universidade Católica de Pelotas, Pelotas, 2004.

AQUINO, C. Interação entre conhecimento explícito e implícito na aprendizagem de L2: quais as contribuições trazidas pela neurolinguística para essa discussão? *Letrônica*, Porto Alegre, v. 5, n. 3, p. 125-141, 2012.

CELCE-MURCIA, M.; BRINTON, D. M.; GOODWIN, J.M; GRINER, B. *Teaching Pronunciation: a course book and reference guide*. Cambridge: Cambridge University Press, 2010.

DEKEYSER, R. M. Más allá de la atención a la forma. In: DOUGHTY, C.; WILLIAMS, J. *Atención a las formas en la adquisición de segundas lenguas en el aula*. España: Edinumen, 2009, p. 57-78.

DIAS, E.C.O. *Declarativas e interrogativas totais no espanhol L1 e L2 falado em Bogotá: uma contribuição para estudos prosódicos*. 2015. 273f. Tese (Doutorado em Linguística) - Universidade Federal de Santa Catarina, Florianópolis. 2015.

ELLIS, R. *La adquisición de segundas lenguas en un contexto de enseñanza análisis de las investigaciones existentes*. Nova Zelandia: ResearchDivision, 2005.

FONTANA, B.; LIMA, M.S. Questões centrais do ensino e da aprendizagem de línguas estrangeiras. *Em Aberto*, Brasília, v.22, n. 81, p. 15-28, ago. 2009.

KRASHEN, S.D. *The Input Hypothesis and its rivals*. London: Longman. 1985.

LONG, M.H. Focus on form: a design feature in language teaching methodology. In: De BOT, K.; GINSBERG, R.; KRAMSCH, C. (Ed.). *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins, 1991.

LYRIO, A. L. L. Ensino comunicativo e foco na forma (FONF). *Litterae - Perspectivas do ensino de línguas e de literaturas*, Vitória, v. 1, n. 1, p. 22-47, 2012.

MARIAN, V.; BLUMENFELD, H. K.; KAUSHANSKAYA, M. The language experience and proficiency questionnaire (leap-q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, Nova Iorque, v. 50, p. 940-967, 2007.

MORENO, M. C. Didáctica de la entonación: una asignatura pendiente. *Dialnet*, Rioja, v. 14, 2002. p. 65-75. Disponível em: < <https://dialnet.unirioja.es/servlet/articulo?codigo=295018>> Acesso em 03 fev.2017.

ORTIZ-PREUSS, E. A interface entre os conhecimentos implícito e explícito: um estudo baseado no *se* – operador aspectual como delimitador em espanhol. 2005. 165f. Dissertação (Mestrado em Letras) – Universidade Católica de Pelotas, Pelotas, 2005.

_____. Similaridade linguística entre português e espanhol: efeitos na produção de fala em L2. *Revista (Con) Textos Linguísticos*, Vitória, v. 8, n. 10, p. 66-82, 2014.

ORTIZ-PREUSS, E.; SANZ, C. Aquisição de L2: Interação entre variáveis externas e internas. In: _____.; COUTO, E. K. N. N. do; RAMOS, R. M. do. (Org.). *Múltiplos olhares e Linguística e Linguística Aplicada*. Campinas: Pontes, 2016, p.121-134.

PINTO, M. S. Transferências prosódicas do PB/LM na aprendizagem do E/LE: enunciados assertivos e interrogativos totais. 2009. 378f. Tese (Doutorado em Letras Neolatinas). Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

PRIETO, P. Teorías lingüísticas de la entonación. In: PRIETO, P. *Teorías de la Entonación*. Barcelona: Ariel, 2003. p.14-33.

Real Academia Española. La entonación. In: _____. *Nueva gramática de la lengua española*. Barcelona: Espasa Libros, 2011. p. 435-488.

SANZ, C. Adult SLA: the interaction between external and internal factors. In: SANZ, C. (Ed.). *Mind and context in adult second language acquisition: methods, theory, and practice*. Washington DC: Georgetown University Press, 2005, p. 3-20.

_____. Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA Project, an appraisal. In: PÉREZ-VIDAL, C. (Ed.). *Language acquisition in study abroad and formal instruction contexts*. Philadelphia/Amsterdam: John Benjamins, 2014. p. 1-13.

SEMINO, M. J. I. La entonación en español y en portugués. In: BRISOLARA, L. B.; SEMINO, M. J. I. *¿Cómo pronunciar el español? La enseñanza de la fonética y la fonología para brasileños: ejercicios prácticos*. Campinas: Pontes Editores, 2014. p.103-126.

SOSA, J.M. *La entonación del español*. Su estructura fónica, variabilidad y dialectología. Madrid: Cátedra, 1999.

SWAIN, M. Three functions of output in second language learning. In: COOK, G.; SEIDLHOFFER, B. (Ed). *Principle & practice in applied linguistics*. Oxford: Oxford University Press, 1995.

VIDAL, R.T. Ensino-aprendizagem do foco na forma: retorno ou recomeço? *TheESpecialist*, São Paulo, v.28, n. 2, p. 159-184, 2007.



Data de submissão: 15/03/2018

Data de aceite: 02/06/2018