Gradus
Revista Brasileira de Fonologia de Laboratório

Detecção automática de fronteiras prosódicas entre unidades entonacionais

Bárbara Helohá Falcão Teixeira (barbaraheloha@gmail.com)
Universidade Federal de Minas Gerais
Tommaso Raso (tommaso.raso@gmail.com)
Universidade Federal de Minas Gerais
Plínio Almeida Barbosa (pabarbosa.unicampbr@gmail.com)
Universidade Estadual de Campinas

Abstract

Speech is segmented into intonational units marked by prosodic boundaries. This work aims both to investigate the phonetic-acoustic parameters that guide the production and perception of prosodic boundaries and to develop models for automatic detection of prosodic boundaries in spontaneous speech. Two samples of male spontaneous speech excerpts were segmented into intonational units by two groups of trained annotators. The boundaries perceived by the annotators were annotated as either terminal (TB) or non-terminal (NTB). A script was used to extract phonetic-acoustic parameters along the speech signal. The extracted parameters comprise measures of: 1) Speech rate and rhythm; 2) Normalized duration; 3) Fundamental frequency; 4) Intensity; 5) Silent pause. A training of models composed by multiple parameters designed to the automatic identification of boundaries marked by the annotators was developed. The Linear Discriminant Analysis algorithm was used and positions at which at least 50% of the annotators indicated a boundary of the same type were considered as boundary. The automatic terminal boundary detection model shows a convergence of 80% in relation to terminal boundaries noticed by annotators in sample I. For non-terminal boundaries, three statistical classification models were obtained. Together, the three models show a convergence of 98% in relation to non-terminal boundaries noticed by annotators in sample I. The models were validated later in sample II. The results of the validation indicate that the performance of the TB model is 74% and that of the NTB model is 88% in sample II.

Article history
Received 2019-09-19
Revised 2020-05-12
Accepted 2020-05-13
Published 2020-08-10

Keywords:
prosodic boundaries
automatic detection
spontaneous speech
speech segmentation

Resumo

A fala é segmentada em unidades entonacionais marcadas por fronteiras prosódicas. Este trabalho tem como objetivo investigar os parâmetros fonético-acústicos que orientam a produção e a percepção de fronteiras prosódicas, bem como desenvolver modelos para detecção automática de fronteiras prosódicas em fala espontânea. Duas amostras de trechos de fala espontânea masculina foram segmentadas em unidades entonacionais por dois grupos de segmentadores treinados. As fronteiras percebidas pelos segmentadores foram anotadas como terminais (TB) ou não-terminais (NTB). Um *script* foi utilizado para extrair parâmetros fonético-acústicos ao longo do sinal de fala. Os parâmetros extraídos compreendem medidas de: 1) Velocidade e ritmo da fala; 2) Duração normalizada; 3) Frequência fundamental; 4) Intensidade; 5) Pausa silenciosa. Foi desenvolvido um treinamento de modelos compostos por múltiplos parâmetros projetados para a identificação automática das fronteiras marcadas pelos segmentadores. Utilizou-se o algoritmo *Linear Discriminant Analysis* e consideraram-se como fronteira posições em que pelo menos 50% dos segmentadores indicaram uma fronteira do mesmo tipo. O modelo de detecção automática de fronteiras terminais mostra uma convergência de 80% em relação às fronteiras terminais observadas pelos segmentadores na Amostra I. Para fronteiras não-terminais, foram obtidos três modelos de classificação estatística. Juntos, os três modelos mostram uma convergência de 98% em relação às fronteiras não-terminais observadas pelos segmentadores na Amostra I. Os modelos foram validados posteriormente na Amostra II. Os resultados da validação indicam que o desempenho do modelo TB é de 74% e o do modelo NTB é de 88% na Amostra II.

Palavras-chave: fronteiras prosódicas; detecção automática; fala espontânea; segmentação da fala.

Introdução

Uma série de estudos linguísticos vêm mostrando que a prosódia exerce uma função fundamental na comunicação oral, transmitindo emoções, construindo diferentes estilos de fala, transmitindo funções comunicativas ilocucionárias ou informacionais, marcando proeminências, desempenhando funções sociolinguísticas, segmentando unidades e estabelecendo as suas respectivas fronteiras prosódicas. No que diz respeito à segmentação da fala em unidades, naturalmente, conforme seja o interesse da pesquisa, há várias unidades de segmentação. Alguns dos exemplos de unidades de segmentação são os fones, as sílabas, os grupos acentuais, as unidades entonacionais, os enunciados, os turnos e até domínios maiores.

Este trabalho visa desenvolver modelos destinados à construção de uma ferramenta automática de detecção de fronteiras prosódicas que demarcam o domínio da unidade entonacional em dados de fala espontânea.[1] A ferramenta funcionará a partir de dois critérios relacionados entre si. Tais critérios são os parâmetros fonético-acústicos extraídos automaticamente do sinal sonoro em conjunto com a percepção de segmentadores treinados para identificar as fronteiras prosódicas. O desenvolvimento deste trabalho tem dois propósitos principais. O primeiro é contribuir ao melhor entendimento teórico acerca da percepção humana das unidades entonacionais e suas fronteiras prosódicas. Do ponto de vista prático, o projeto visa a desenvolver uma ferramenta computacional que auxilie na compilação de corpora de fala espontânea do português do Brasil (doravante PB), tornando o processo de segmentação da fala mais rápido, poupando-se simultaneamente tempo e esforços humanos, o que pode ser visto como uma contribuição para a linguística de *corpus* em geral.

[1] Como fala espontânea entendemos aqui as instâncias de comunicação humana não planejadas em contexto natural ou em contextos midiáticos, sem intervenção do pesquisador, abrangendo interações dialógicas, monológicas e conversacionais.

A segmentação da fala em unidades

A fala é realizada e percebida em pequenos agrupamentos de uma ou poucas palavras, também chamados de unidades entonacionais.[2] A definição de unidade entonacional abrange várias concepções a depender da teoria. Em geral, uma unidade entonacional pode ser definida como um grupo de palavras delimitado por meio de fronteiras prosódicas relevantes perceptualmente e marcadas fisicamente no sinal acústico através de um contorno entonacional coerente, diferente dos contornos precedentes e sub-

[2] Schubiger, *English intonation* (1958); Chafe, "The deployment of consciousness in the production of a narrative" (1980); Schuetze-Coburn, "Prosody, syntax, and discourse pragmatics" (1994); Ladd, *Intonational Phonology* (2008).

sequentes.[3] Essa definição não é satisfatória, porque falta uma boa definição de fronteira que não gere circularidade. Também, falta uma definição de "contorno entonacional coerente", assim como sabemos que nem sempre a entonação é um elemento decisivo ou até necessário para a percepção e a marcação de uma fronteira entre duas unidades. No âmbito formalista[4] e no sistema de anotação ToBI,[5] a marcação de fronteira é identificada em alguns parâmetros formais baseados no tom e na duração.

A segmentação da fala em unidades entonacionais e suas fronteiras prosódicas têm sido consistentemente exploradas no âmbito de compilação de grandes *corpora* orais de PB, inglês, italiano, francês, espanhol, hebraico, português europeu, holandês, etc. Para o inglês, podemos citar o *Santa Barbara Corpus of Spoken American English*,[6] o *corpus AixMARSEC*,[7] o *Hong Kong Corpus of Spoken English*[8] e o *Corpus News* da Rádio da Universidade de Boston;[9] para português europeu, italiano, francês e espanhol, foram publicados os *corpora* da família C-ORAL-ROM;[10] para o português brasileiro, o *corpus* C-ORAL-BRASIL I [11] de fala informal, e está prestes a ser publicado o C-ORAL-BRASIL II, de fala formal em contexto natural, mídia e telefone;[12] para o hebraico, podemos citar o *Corpus of Spoken Israeli Hebrew*;[13] para o holandês, temos o *Spoken Dutch Corpus*;[14] para línguas diversas, podemos citar o *Corpus of Spoken AfroAsiatic Languages*.[15]

As unidades entonacionais podem ser analisadas funcionalmente segundo perspectivas teóricas diferentes: sintáticas,[16] pragmáticas[17] e cognitivas.[18] No entanto, as fronteiras podem ser estudadas *per se*, independentemente da perspectiva teórico-funcional que orienta a interpretação da unidade entonacional, porque as fronteiras desempenham claramente um papel na compreensão dos textos falados, pois compreendem um fenômeno atestado perceptualmente.

Neste contexto, uma questão central que surge é: qual tipo de informação as fronteiras fornecem? A pergunta pode ser respondida por diversos autores da área. Estudos evidenciam que as fronteiras de natureza prosódica são utilizadas pelo falante para guiar o ouvinte na reconstrução da segmentação pretendida e na compreensão adequada da mensagem.[19] Conforme é observado no exemplo abaixo, as fronteiras têm como função captar adequadamente o domínio das relações linguísticas em textos falados.

Exemplo – João vai pro Rio até amanhã.

i. João vai pro Rio até amanhã (asserção).

ii. João (chamamento)! Vai pro Rio até amanhã (ordem)!

iii. João (chamamento)! Vai pro Rio (ordem)! Até amanhã (despedida).

3 du Bois et al., "Discourse transcription" (1992); Cruttenden, *Intonation* (1997).

4 Pierrehumbert, "The phonetics and phonology of English intonation" (1980); Pierrehumbert et al., "Conceptual foundations of phonology as a laboratory science" (2000).

5 Silverman et al., "ToBI: A standard for labeling English prosody" (1992).

6 du Bois et al., *Santa Barbara corpus of spoken American English* (2000).

7 Auran et al., "The Aix-MARSEC project: an evolutive database of spoken British English" (2004).

8 Cheng et al., "The creation of a prosodically transcribed intercultural corpus" (2005).

9 Ostendorf et al., *The Boston University radio news corpus* (1995).

10 Cresti e Moneglia, *C-ORAL-ROM* (2005).

11 Raso e Mello, *C-ORAL-BRASIL I* (2012).

12 Raso et al., *C-ORAL-BRASIL II* (sem data).

13 `http://cosih.com/english/index.html`

14 Schuurman et al., "CGN, an annotated corpus of spoken Dutch" (2003).

15 Mettouchi e Chanard, "From fieldwork to annotated corpora" (2010).

16 Cooper e Paccia-Cooper, *Syntax and speech* (1980); Selkirk, "Comments on intonational phrasing in English" (2005).

17 Halliday, "Speech and situation" (1965); Cresti, *Corpus di italiano parlato* (2000); Reed, "Prosody, syntax and action formation" (2012).

18 Chafe, *Discourse, consciousness, and time* (1994); Croft, "Intonation units and grammatical structure" (1995); Bybee, *Language, usage and cognition* (2010).

19 Swerts, "Prosodic features at discourse boundaries of different strength" (1997); Watson e Gibson, "The relationship between intonational phrasing and syntactic structure in language production" (2004); Frazier et al., "Prosodic phrasing is central to language comprehension" (2006).

iv. João (pedido de confirmação)? Vai pro Rio até amanhã (resposta).

v. João vai pro Rio (pedido de confirmação)? Até amanhã (despedida).

No exemplo descrito acima, a estrutura pode ser segmentada em uma única unidade, duas ou três unidades, mas, naturalmente, a estrutura também pode ser segmentada em uma maior quantidade de unidades. Independentemente das ilocuções que podem ser veiculadas nas unidades em que o trecho foi segmentado, as fronteiras prosódicas fornecem informações morfossintáticas cruciais para a interpretação dos textos falados. Com a opção (i), por exemplo, *João* se configura como sujeito do verbo *vai*, que por sua vez será terceira pessoa do presente indicativo. Em (ii) e em (iii) *João* não pode ser considerado o sujeito do verbo *vai*, mas sim um enunciado autônomo, e *vai* deve ser analisado como segunda pessoa do imperativo. *Até amanhã* pode ser considerado um adjunto em (i), em (ii) e em (iv), mas em (iii) e (v) constitui por si só um enunciado totalmente autônomo cujo significado é muito diferente daquele dado pelos adjuntos.

As fronteiras prosódicas também podem desempenhar um papel na sinalização da presença de dependências de longa distância[20] e podem resolver casos de ambiguidade local no discurso falado.[21] Assim, de modo geral, as fronteiras prosódicas podem ser vistas como índices linguísticos de suma importância para depreender o significado nos níveis pragmático, semântico e sintático da comunicação falada.

As perspectivas perceptuais caracterizam e descrevem as fronteiras prosódicas de diferentes formas. Alguns trabalhos optam por investigar a oposição entre presença *versus* ausência de fronteira não diferenciando entre si as fronteiras percebidas.[22] Outros estudos optam por uma distinção mais ampla em que são estabelecidos diferentes tipos de fronteiras prosódicas. Tal distinção, no entanto, não é comum aos diversos trabalhos.

Dentre os trabalhos que estabelecem algum tipo de distinção entre as fronteiras perceptualmente relevantes, alguns argumentam que as fronteiras prosódicas não devem ser consideradas um fenômeno linguístico categórico, mas sim um fenômeno gradiente no qual são estabelecidos níveis de força. Neste caso, considera-se que as fronteiras são um fenômeno gradiente organizado gradativamente por meio de diferentes níveis de força limitados.

Porém, dentre os diferentes autores que optam por essa alternativa, existe uma clara discordância sobre os possíveis níveis de força pelos quais as fronteiras seriam produzidas e percebidas.[23] Desta forma, ainda há muitas discussões sobre a quantidade de fronteiras que são produzidas e percebidas com níveis de força di-

[20] Kraljic e Brennan, "Prosodic disambiguation of syntactic structure" (2005); Schafer et al., "Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task" (2005); Snedeker e Trueswell, "Using prosody to avoid ambiguity" (2003).

[21] Kjelgaard e Speer, "Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity" (1999); Speer et al., "The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities" (1996); Warren et al., "Prosody, phonology and parsing in closure ambiguities" (1995).

[22] Mo et al., "Naïve listeners' prominence and boundary perception" (2008); Barbosa, "Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech" (2010); de Pijper e Sanderman, "On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues" (1994).

[23] Para uma discussão completa sobre esse recorrente desacordo, veja-se Barbosa, *Incursões em torno do ritmo da fala* (2006).

ferentes. Alguns autores distinguem três níveis de força[24], alguns estudos individualizam mais de dois níveis.[25]

Também dentro das abordagens perceptuais, alguns autores propõem a existência de um número determinado de fronteiras. Neste caso, argumenta-se que os tipos de fronteiras são associados à percepção de conclusão ou de continuação da unidade.[26] De forma geral, esses trabalhos têm uma visão categórica sobre o fenômeno e a nomenclatura adotada estabelece que as fronteiras que indicam perceptualmente a continuação (não-conclusão) da unidade entonacional são chamadas de fronteiras não-terminais (NTB), já as fronteiras que indicam a finalização da unidade são chamadas de fronteiras terminais (TB).

Independentemente da perspectiva de análise, argumenta-se que as fronteiras prosódicas são marcadas no sinal acústico por meio de certas configurações de parâmetros fonético-acústicos. No entanto, estas configurações ainda não são completamente compreendidas. Por um lado, essa falta de compreensão pode ser justificada por questões relacionadas à metodologia empregada; por outro, ela também se justifica por um fator muito relevante: o recorrente desacordo entre as hipóteses a respeito da quantidade e dos tipos de fronteiras prosódicas estabelecidas. Os trabalhos que analisam o estabelecimento das fronteiras do ponto de vista da percepção ainda enfrentam dificuldades metodológicas para a análise de um aspecto tão complexo como a percepção humana. Para um apanhado geral sobre os problemas fonéticos, cognitivos e metodológicos relativos aos estudos das fronteiras prosódicas, vejam-se os trabalhos listados.[27]

Este trabalho insere-se dentro da perspectiva de estudo que visa a compreender as fronteiras de unidades entonacionais de um ponto de vista perceptual e fonético-acústico. Optou-se pela adoção de unidades entonacionais percebidas como conclusivas ou não-conclusivas e pela hipótese de que haja uma diferença clara entre fronteiras terminais e não terminais. Portanto, o trabalho não deve ser visto como ligado a uma teoria especificamente, mas simplesmente como um trabalho que parte da hipótese de uma distinção funcional entre dois tipos de fronteiras perceptivelmente relevantes chamadas de fronteiras terminais e não-terminais.

Ferramentas de segmentação automática da fala

De fato, o desenvolvimento tecnológico de ferramentas para análise e extração de dados acústicos possibilitou o desenvolvimento de ferramentas de detecção automática de fronteiras prosódicas. Uma revisão de literatura da área mostra que há uma série de ferramentas construídas com o objetivo de identificar fronteiras

[24] Mertens e Simon, "Towards Automatic detection of prosodic boundaries in spoken French" (2013).

[25] Wightman et al., "Segmental durations in the vicinity of prosodic phrase boundaries" (1992), para o inglês; Barbosa, "Caractérisation et génération automatique de la structuration rythmique du français" (1994); Barbosa, *Incursões em torno do ritmo da fala* (2006).

[26] Pike, *The Intonation of American English* (1945); Moneglia e Cresti, "L' intonazione e i criteri di trascrizione del parlato adulto e infantile" (1997); Swerts, "Prosodic Features of Discourse Units" (1994).

[27] Barth-Weingarten, *Intonation Units Revisited* (2016); Izre'el et al., *In Search of Basic Units of Spoken Language* (2020).

prosódicas classificadas de acordo com os índices adotados no sistema notacional ToBI.[28]

É importante salientar também que muitas dessas ferramentas foram construídas com base em dados de fala lida ou fala exclusivamente radiofônica. As ferramentas construídas com base nestes dados, em sua maioria, indicam apenas a presença de fronteira, não informando o índice da fronteira. Em outras palavras, tais ferramentas não distinguem entre os diferentes tipos de fronteiras propostos no sistema notacional ToBi.

Certamente, as ferramentas citadas oferecem uma boa contribuição a respeito da segmentação automática da fala em unidades, porque foram utilizados vários métodos de análise estatística e diversos tipos de variáveis que podem ajudar no desenvolvimento de novas ferramentas. No entanto, tais ferramentas não oferecem um sistema de detecção automática de fronteiras prosódicas que são relevantes perceptualmente em situações de comunicação real entre falantes (fala espontânea). Além disso, tais ferramentas não são capazes de indicar o valor de continuidade ou terminalidade associado a uma fronteira. Assim, no âmbito das ferramentas de segmentação automática da fala, é necessário que outras ferramentas sejam desenvolvidas com o intuito de identificar automaticamente fronteiras prosódicas relevantes perceptualmente na fala espontânea, pois, até o momento, a área não dispõe de uma ferramenta com essa finalidade.

[28] Silverman et al., "ToBI: A standard for labeling English prosody" (1992); Wightman e Ostendorf, "Automatic labeling of prosodic patterns" (1994); Ross e Ostendorf, "Prediction of abstract prosodic labels for speech synthesis" (1996); Ni et al., "Automatic prosodic break detection and feature analysis" (2012); Anantha-krishnan e Narayanan, "An Automatic prosody recognizer using a coupled multi-stream acoustic model and a syntactic-prosodic language model" (2005).

Objetivos

i. Descrever e analisar os parâmetros fonético-acústicos que orientam a produção e a percepção de dois tipos de fronteiras prosódicas na fala espontânea;

ii. Uma vez identificados os parâmetros associados às macro-categorias de fronteiras, buscar um refinamento interno, sempre em correspondência com a percepção humana;

iii. Desenvolver modelos de classificação estatística para predizer a realização de fronteiras prosódicas em dados de fala espontânea em PB;

iv. Contribuir à busca de um processo mais rápido de compilação de *corpora* orais segmentados em unidades entonacionais.

Metodologia

Dados

Os dados compreendem quatorze trechos de fala monológica masculina com alta qualidade acústica (mínimo ruído de fundo e ausência de sobreposição de fala). Os dados foram extraídos tanto da parte informal, quanto da parte formal e de mídia do *corpus* C-ORAL-BRASIL I[29] e II.[30] Os textos são, portanto, relativos a três *corpora* de fala espontânea: fala informal em contexto natural, fala formal em contexto natural e fala televisiva. Os trechos são compostos por em média 192 palavras e são distribuídos da seguinte forma em duas amostras:

[29] Raso e Mello, *C-ORAL-BRASIL I* (2012).
[30] Raso et al., *C-ORAL-BRASIL II* (sem data).

Tabela 1: Descrição dos dados.

Contexto	Amostra	Texto	Duração	Palavras
Natural informal	I	bfammn11	01'11''	189
		bfammn24	00'58''	151
	II	bpubmn12	01'26''	198
		bpubmn13	01'00''	180
Mídia formal	I	bmidmasc01	01'23''	212
		bmidmasc02	01'21''	238
		bmidmasc03	01'07''	183
	II	bmedsp03_1a 0	1'02'' 2	06
		bmedsp03_1b 0	1'07'' 2	00
		bmedts10_1 0	1'11'' 1	80
Natural formal	I	bnatmasc01	01'30''	205
		bnatmasc02	01'09''	161
	II	bnatc003	01'00''	202
		bnatpr05	01'43''	181

Tratamento dos dados

O tratamento dos dados deste trabalho foi dividido em duas etapas. Na primeira etapa, cada trecho foi segmentado autonomamente por dois grupos de segmentadores *experts*, membros da equipe do Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Universidade Federal de Minas Gerais (UFMG). O grupo que segmentou a Amostra I é composto por 14 pessoas, já o grupo que realizou a segmentação da Amostra II é composto por 19 pessoas. Todos os segmentadores foram treinados anteriormente pela equipe do laboratório e já apresentavam,

mesmo que em diferentes graus, experiência em segmentação prosódica da fala.

Tendo em vista os critérios de segmentação previstos na compilação dos *corpora* C-ORAL-BRASIL, além das fronteiras terminais e não-terminais, estão previstas as marcações de duas outras fronteiras prosódicas denominadas *retractings* e interrupções. Os *retractings* são unidades retratadas pelos falantes e são marcadas por fronteiras não-terminais; as interrupções compreendem unidades entonacionais em que o falante interrompe o enunciado por motivos próprios ou por fatores externos, e são marcadas por quebras terminais. Em ambos os casos se mantém o registro da diferença desses casos com os outros tipos de quebras. Assim sendo, os segmentadores foram apresentados às gravações e aos textos transcritos, sem nenhuma marcação de fronteiras, e tinham que realizar a tarefa de segmentação prosódica.

A tarefa dos segmentadores consistia em marcar as posições em que as fronteiras prosódicas foram percebidas, inserindo na transcrição recebida uma barra simples (/) para fronteira não-terminal, uma barra dupla (//) para fronteira terminal, uma barra simples entre colchetes para *retractings* ([/]) e um sinal de adição para interrupções (+). A marcação de *retractings* e interrupções é necessária para minimizar erros na detecção automática de fronteiras prosódicas decorrentes de fronteiras que podem ser consideradas disfluências e não são propriamente fronteiras produzidas por razões linguísticas, ou seja, não respondem a uma estratégia planejada.

Na segunda etapa, os trechos foram anotados em sete camadas usando o *software* para análises fonéticas e acústicas PRAAT.[31] Foram adotadas as seguintes camadas de anotação do TextGrid do PRAAT:

[31] https://www.praat.org/.

i. Segmentação em unidades Vogal-Vogal (unidades V-Vs)[32] e etiquetagem[33] utilizando transcrição fonética larga em caracteres ASCII;

ii. Anotação das fronteiras não-terminais marcadas pelos segmentadores, informando o número de pessoas que marcaram a fronteira;

iii. Anotação das fronteiras terminais marcadas pelos segmentadores, informando o número de pessoas que marcaram a fronteira;

iv. Anotação de *retractings* marcados pelos segmentadores, informando o número de pessoas que marcaram o *retracting*;

v. Anotação de interrupções marcadas pelos segmentadores, informando o número de segmentadores que marcaram a interrupção;

[32] A segmentação em unidades V-Vs foi feita automaticamente utilizando-se o *script* BEATEXTRACTOR para PRAAT, desenvolvido pelo terceiro autor. Posteriormente, foi realizada uma revisão manual da segmentação gerada. A revisão foi realizada por duas pessoas treinadas para realizar esse tipo de segmentação.

[33] A etiquetagem foi realizada manualmente por duas pessoas treinadas para realizar esse tipo de etiquetagem.

vi. Anotação do intervalo referente a pausas silenciosas;

vii. Transcrição ortográfica do texto do enunciado.

Abaixo, a fig. 1 mostra um exemplo das camadas de anotação adotadas. O exemplo mostra as primeiras unidades entonacionais marcadas pelo grupo de segmentadores no trecho BMEDSP03_1B.

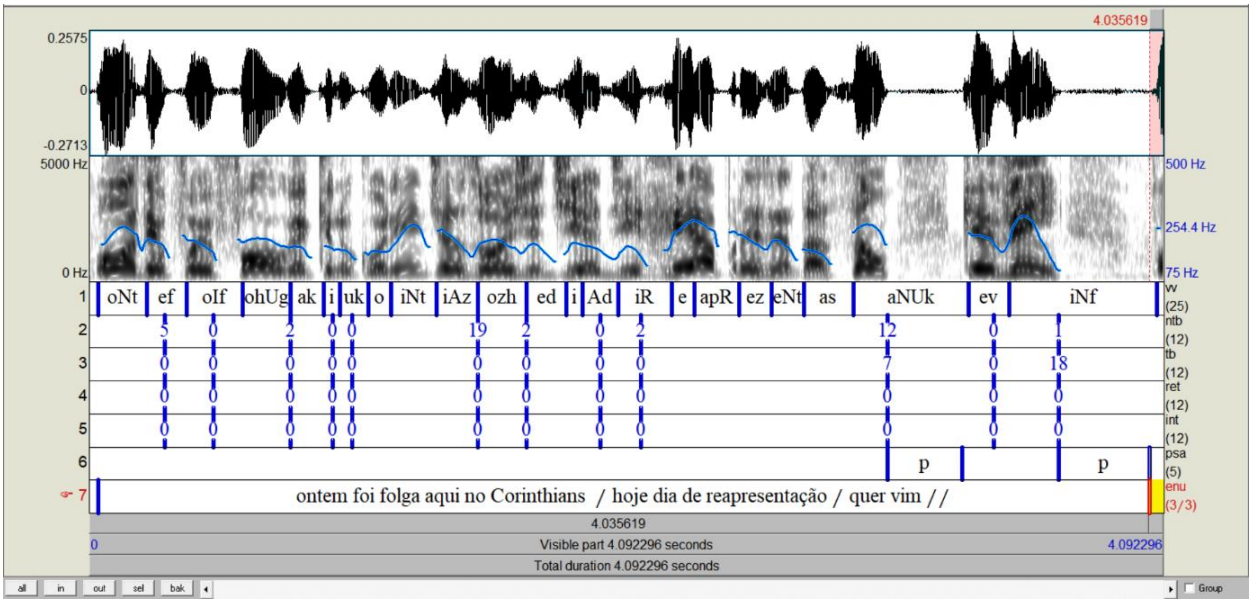

Figura 1: Camadas de anotação do PRAAT. Trecho: ontem foi folga aqui no Corinthians / hoje dia de reapresentação / quer vim //

Com o intuito de analisar todas as potenciais posições de estabelecimento de fronteira prosódica, foi utilizado o *script* BREAKDESCRIPTOR v2.0,[34] para PRAAT, para extrair uma série de parâmetros fonético-acústicos em todas as unidades V-Vs em uma janela centrada em toda fronteira de palavra fonológica. Isso inclui naturalmente as posições indicadas pelos segmentadores como posições de fronteiras prosódicas, mas inclui também as posições indicadas pelos segmentadores como não fronteiras. O *script* extrai as medidas inclusive nos casos de não fronteiras, porque o conjunto de parâmetros associados às não fronteiras são fundamentais para a classificação estatística.

[34] Desenvolvido pelo terceiro autor.

O funcionamento do BREAKDESCRIPTOR v2.0 é estabelecido a partir de dois critérios de referência. O primeiro critério consiste em definir a extensão da janela de trabalho a ser analisada pelo *script*; o segundo compreende a porcentagem de acordo entre segmentadores a ser considerada para estabelecer quais são os pontos de fronteiras e não fronteiras. Deste modo, a janela de análise é móvel, podendo ser composta por até 10 unidades anteriores e 10 unidades posteriores em relação à posição de análise. Em relação ao acordo entre segmentadores, o *script* permite que seja atribuído qualquer porcentagem de referência.

Neste trabalho, optou-se pela adoção do acordo de pelo menos

50% entre segmentadores para estabelecer as posições de fronteiras. Na Amostra I, as posições de fronteira prosódica são aquelas em que pelo menos 7 segmentadores marcaram necessariamente uma fronteira do mesmo tipo; já na Amostra II, as posições de fronteira são aquelas em que pelo menos 10 segmentadores marcaram uma fronteira da mesma natureza. Optou-se pela adoção de uma janela de trabalho composta por 10 unidades anteriores e posteriores em relação à posição analisada. No total, para cada unidade V-V localizada em fronteira de palavra fonológica, o *script* analisa e extrai os parâmetros em 21 unidades (10 anteriores, 10 posteriores e a própria unidade em análise). Abaixo, a fig. 2 mostra as janelas de trabalho do *script*.

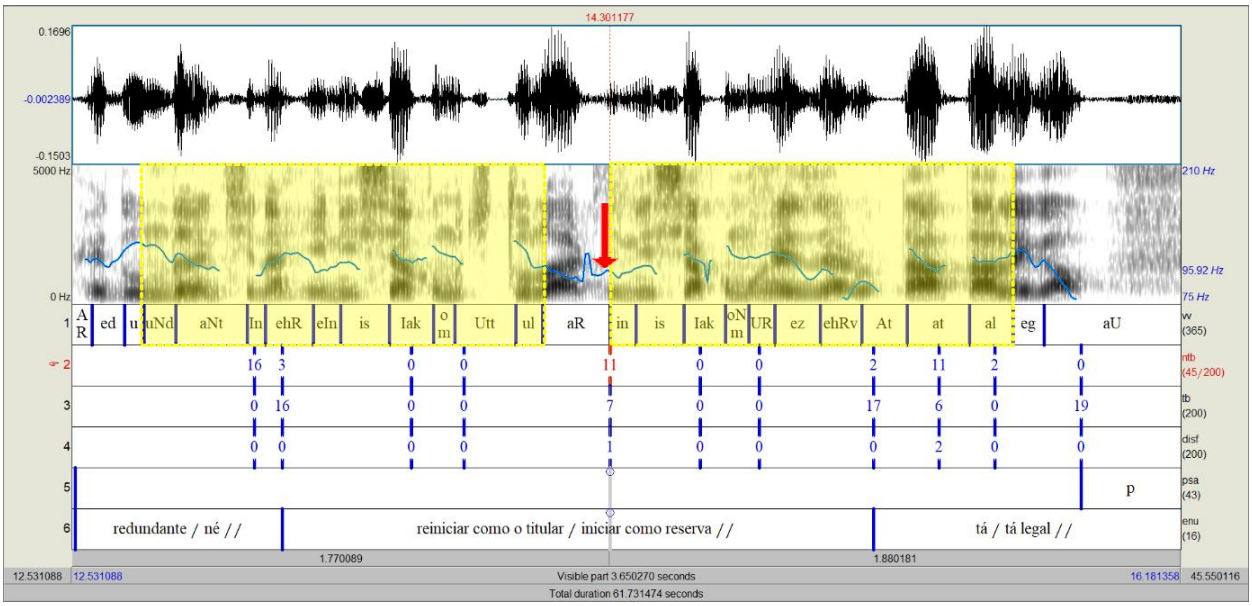

Figura 2: Funcionamento do *script* BreakDescriptor v2.0, para Praat. A fronteira destacada está inserida no enunciado: reiniciar como titular / iniciar como reserva //

Na fig. 2, no centro, destaca-se uma unidade V-V que 11 segmentadores marcaram como posição de NTB, 7 marcaram como TB e um marcou como disfluência. Em virtude da porcentagem de acordo adotada para estabelecer os diferentes tipos de fronteira prosódica, a fronteira destacada é considerada uma NTB. Para a posição destacada com a seta vermelha, são calculados os parâmetros fonético-acústicos nas 10 unidades V-Vs anteriores (área sombreada do espectrograma), na unidade que marca a posição da fronteira e nas 10 unidades V-Vs posteriores à posição de análise (área sombreada do espectrograma).

Apesar de os segmentadores terem marcado posições em que perceberam *retractings* e interrupções, no atual momento de desenvolvimento deste trabalho, optou-se pela não utilização dessas posições, porque tais fronteiras parecem ser involuntárias, configurando-se como disfluências pouco padronizadas no que diz respeito às configurações de parâmetros fonético-acústicos; funcionariam portanto como elementos de confusão na busca de

padrões. A seguir, na tabela 2, em conformidade com a porcentagem de acordo entre segmentadores estabelecida para determinar as categorias associadas às posições (no mínimo 50%), explicitamos a composição das duas amostras.

Tabela 2: Posições analisadas.

Categoria	Total	%	Amostra	Frequência
Fronteira terminal	116	4,8	I	70
			II	46
Fronteira não-terminal	534	22,3	I	242
			II	292
Nenhuma fronteira	1744	72,8	I	985
			II	759

A tabela 2 mostra que, no total, foram analisadas 2394 posições. Os dados contêm 116 posições de TB, 534 de NTB e 1744 posições de nenhuma fronteira prosódica (NB).

Independentemente da extensão da janela e porcentagem de acordo entre os segmentadores para definir as posições de presença e ausência de fronteira, as medidas extraídas automaticamente pelo *script* adotado sempre compreendem medidas de natureza global e local. No que tange às medidas extraídas, a única diferença é a quantidade de medidas locais extraídas. Assim, sempre são extraídas 42 medidas globais. A quantidade de medidas locais extraídas é diversificada e proporcionalmente dependente da extensão de janela escolhida, de modo que quanto maior for a extensão da janela, maior é a quantidade de medidas locais extraídas. Para a maior extensão de janela que pode ser atribuída (10 unidades V-Vs anteriores, 10 unidades V-Vs posteriores e a própria unidade V-V em análise), no total, são extraídas pelo *script* 111 medidas (42 globais e 69 locais) que permitem desenvolver análises de fenômenos prosódicos estabelecidos ao longo do sinal acústico.

As propriedades físicas relativas à produção dos sons extraídas automaticamente pelo *script* BreakDescriptor compreendem medidas de:

i. Taxa de elocução e ritmo (6 medidas)
ii. Duração normalizada dos segmentos silábicos[35] (no máximo, 34 medidas)
iii. Frequência fundamental (no máximo, 65 medidas)
iv. Intensidade (4 medidas)
v. Pausa (2 medidas).

[35] As referidas medidas são obtidas automaticamente pelo *script* por meio de um procedimento de normalização básico em estatística, que consiste em obter a duração normalizada das unidades V-Vs (z–score estatístico que indica o afastamento do valor medido em relação a uma média em unidades de desvio-padrão). Em seguida, as curvas das durações normalizadas são suavizadas por meio de um procedimento padrão de média ponderada.

A seguir, a tabela 3 mostra, de forma sintética, os parâmetros extraídos automaticamente pelo *script*.[36]

[36] O detalhamento das medidas extraídas pode ser encontrado em Teixeira, "Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea" (2018).

Grupo	Tipo	Medidas
Taxa de elocução (*speech rate*) e ritmo	Global	Taxa de unidades V-V por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Taxa de unidades V-V não-salientes por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Duração normalizada dos segmentos silábicos	Local	Média de z-score suavizado de duração da unidade V-V (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre o z-score de duração da unidade V-V imediatamente após a posição de análise e o z-score de duração da unidade V-V analisada)
	Global	Média de z-score suavizado de duração da unidade V-V (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Desvio-padrão de z-score suavizado de duração das unidades V-Vs (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Assimetria de z-score suavizado de duração das unidades V-Vs (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Taxa de picos de z-score suavizado de duração das unidades V-Vs por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Frequência fundamental	Local	Mediana de F_0 em semitons re 1 Hz (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre a mediana de F_0 da unidade V-V imediatamente após a posição de análise e mediana de F_0 da unidade V-V analisada) Primeira derivada da mediana de F_0 em semitons re 1 Hz (em cada unidade V-V que compõe as janelas à direita e à esquerda; diferença entre a primeira derivada da mediana de F_0 da unidade V-V imediatamente após a posição de análise e a primeira derivada da mediana de F_0 da unidade V-V analisada)
	Global	Média das medianas de F_0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Desvio-padrão das medianas de F_0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Assimetria das medianas de F_0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Média das primeiras derivadas das medianas de F_0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Desvio-padrão das primeiras derivadas das medianas de F_0 em semitons re 1 Hz (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos) Taxa de picos de F_0 suavizado por segundo (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Intensidade	Local	Média de ênfase espectral em dB (diferença entre a média de ênfase espectral da unidade V-V imediatamente após a posição de análise e a média de ênfase espectral da unidade V-V analisada)
	Global	Média de ênfase espectral em dB (contexto de janela à direita, contexto de janela à esquerda e a diferença entre os dois contextos)
Pausa	Local	Presença de pausa silenciosa (0 = ausência ou 1 = presença) Duração da pausa silenciosa em segundos

Tabela 3: Parâmetros fonético-acústicos extraídos automaticamente pelo *script* BreakDescriptor para Praat.

Análise estatística

Algoritmos de Aprendizagem de Máquina (em inglês, *Machine Learning*) são sistemas computacionais analíticos com alta capacidade de aprender características relevantes sobre diversos fenômenos a partir da análise de dados. O conhecimento aprendido pela máquina tem como objetivo prever, em outros dados, a realização do fenômeno em questão. A principal vantagem do uso destes algoritmos é o desenvolvimento de modelos automáticos, que analisam uma grande quantidade de dados e executam tarefas rapidamente de forma muito precisa. Um conceito importante na área de *Machine Learning* é a classificação estatística. A classificação estatística pode ser entendida como o processo em que etiquetas (categorias) são atribuídas automaticamente por meio de algoritmos de *Machine Learning* treinados para este fim.

Tendo em vista os objetivos deste trabalho e o tratamento recebido pelos dados, foi utilizado o algoritmo supervisionado *Linear Discriminant Analysis* (LDA) para realizar a classificação das fronteiras prosódicas. O processo de treinamento dos modelos consistiu em construir modelos compostos por múltiplos parâmetros fonético-acústicos extraídos automaticamente pelo BreakDescriptor. A construção dos modelos foi feita heuristicamente e todas as medidas extraídas do BreakDescriptor foram avaliadas. Deste modo, avaliou-se o efeito da inclusão ou exclusão de cada uma das medidas no desempenho alcançado pelos modelos.[37]

[37] O detalhamento do processo de treinamento dos modelos pode ser encontrado em Teixeira, "Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea" (2018).

No treinamento destinado às fronteiras terminais, foram usadas as posições de TB, NTB e nenhuma fronteira. No caso do modelo TB, foram usadas as categorias presença de TB e ausência de TB (NO-TB). A categoria engloba posições de TB, a categoria NO-TB engloba posições de NTB e nenhuma fronteira (NB).

Devido a uma maior dificuldade do algoritmo para detectar as fronteiras não-terminais, no treinamento destinado a essas fronteiras, foram usadas as posições de NTB e nenhuma fronteira, ou seja, foram excluídas as posições TB. Deste modo, foram usadas as categorias NTB e NB. O procedimento de treinamento voltado para as fronteiras não-terminais também consistiu em treinar o modelo heuristicamente, eliminar as fronteiras não-terminais identificadas pelo modelo que obteve o melhor desempenho na identificação automática das fronteiras não-terminais e desenvolver outro processo de treinamento heurístico para detectar as fronteiras remanescentes não classificadas corretamente, quantas vezes fosse necessário refazer este processo.

A análise estatística dos dados foi realizada por meio do ambiente para computação estatística R.[38] Os dados discriminados na Amostra I foram diretamente submetidos ao treinamento do

[38] https://www.R-project.org/.

algoritmo adotado. Os dados referentes à Amostra II, por sua vez, foram utilizados no teste final dos modelos desenvolvidos através da Amostra I. Assim, de modo geral, as Amostras I e II foram usadas respectivamente no desenvolvimento e validação dos modelos.

Após a realização da validação dos modelos por meio dos dados dispostos na Amostra II, todos os dados foram combinados em uma única base de dados que contém todos os trechos analisados para que os modelos obtidos fossem reaplicados na base de dados completa. Adicionalmente, foram observados os diferentes tipos de acordo e desacordo entre segmentação realizada pelos segmentadores e segmentação prevista pelos modelos, utilizando-se a base completa de dados.

Resultados

Performance dos modelos

Foram obtidos quatro modelos de detecção automática de fronteiras prosódicas durante o treinamento realizado com os dados da Amostra I. Um deles é destinado à detecção de TB; os demais dedicam-se à detecção de NTB. Abaixo, apresentamos a relevância hierárquica dos parâmetros incluídos em cada um dos modelos e o desempenho alcançado pelos modelos nas fases de desenvolvimento e validação deles.

				Frequência de fronteiras identificadas	
Modelo	**Principais parâmetros**	**Amostra**	**Etapa**	**n**	**%**
TB	Pausa e F_0	I	Desenvolvimento	56	80
		II	Validação	34	74
NTB 1	Duração e pausa	I	Desenvolvimento	152	68
		II	Validação	193	66
NTB 2	Taxa de elocução e F_0	I	Desenvolvimento	57	25
		II	Validação	55	19
NTB 3	Duração e F_0	I	Desenvolvimento	11	5
		II	Validação	10	3

Tabela 4: Modelos de detecção automática de fronteiras.

A reaplicação dos modelos na base completa de dados apresentou os seguintes resultados dispostos nas tabelas 5 e 6. As situações fundamentais de acordo entre os segmentadores e os modelos TB, NTB 1, NTB 2 e NTB 3 estão expostas na tabela 5.

Desempenho	Marcação do grupo	Predição do modelo	Modelo	Frequência	%
Acordo	TB	TB	TB	90	77.6
Acordo	NTB	NO-TB	TB	352	65.9
Acordo	NB	NO-TB	TB	1680	96.3
Acordo	NTB	NTB	NTB 1	351	65.7
			NTB 2	213	39.9
			NTB 3	121	22.7
Acordo	NB	NB	NTB 1	1289	73.9
			NTB 2	350	20
			NTB 3	86	4.9

Tabela 5: Acordos entre segmentadores e modelos. Base completa de dados.

Abaixo, as situações de desacordo entre segmentadores e modelos são expostas a seguir na tabela 6.

Desempenho	Marcação do grupo	Predição do modelo	Modelo	Frequência	%
Desacordo	TB	NO-TB	TB	25	21.6
Desacordo	NTB	TB	TB	162	30.3
Desacordo	NB	TB	TB	5	0.3
Desacordo	TB	NTB	NTB 1	101	87.1
			NTB 2	32	27.6
			NTB 3	28	24.1
Desacordo	TB	NB	NTB 1	15	12.9
			NTB 2	73	62.9
			NTB 3	88	75.9
Desacordo	NTB	NB	NTB 1	181	33.9
			NTB 2	289	54.1
			NTB 3	412	77.2
Desacordo	NB	NTB	NTB 1	447	25.6
			NTB 2	648	37.2
			NTB 3	287	16.5

Tabela 6: Desacordos entre segmentadores e modelos. Base completa de dados.

Composição dos modelos

Apresentamos a seguir a composição dos modelos de identificação de fronteiras prosódicas. Primeiramente, apresentamos o modelo de identificação de fronteiras terminais.

Rank	Sigla	Fenômeno	Peso
1º	psdur	duração da pausa após a fronteira	2.641 517
2º	psp	presença de pausa após a fronteira	1.948 023
3º	fomeddloc	*reset* de F_0	0.328 589
4º	dfomedr1	mudança no contorno entonacional de F_0 da unidade pós-fronteiriça	0.264 072
5º	dfomedl	tendência geral do contorno entonacional de F_0 na janela anterior à fronteira	0.257 113
6º	sddfod	mudança na regularidade do contorno entonacional de F_0 – medida através da primeira derivada da mediana de F_0	0.157 061
7º	prd	mudança na taxa de saliência duracional	0.100 95
8º	sdfol	variação da F_0 média na janela anterior à fronteira	0.091 317
9º	dfomedl10	mudança no contorno entonacional de F_0 da unidade pré-fronteiriça	0.065 812
10º	forl	taxa de proeminência de F_0 (picos de F_0 por segundo) na janela anterior à fronteira	0.032 949
11º	dfomeddloc	mudança no contorno de F_0 entre a unidade pós-fronteiriça (R1) e fronteiriça (L0) – medida a partir da primeira derivada da mediana de F_0	0.032 238
12º	fomedd	mudança na F_0 média da janela posterior à fronteira em relação à F_0 média da janela anterior à fronteira	0.029 097
13º	zl10	duração normalizada da unidade imediatamente anterior à fronteira	0.027 885
14º	skfod	mudança na predominância do contorno entonacional (de crescente para decrescente ou vice-versa) - medido através da média	0.025 454
15º	mzd	mudança na duração da janela posterior à fronteira em relação à duração da anterior à fronteira	0.014 874
16º	skdfod	mudança na predominância do contorno entonacional (de crescente para decrescente ou vice-versa) - medido através da primeira derivada da mediana de F_0	0.010 724
17º	SDzl	variação no ritmo na janela anterior à fronteira	0.009 925
18º	ard	mudança na taxa de articulação	0.003 481
19º	zdloc	mudança na duração normalizada da unidade pós-fronteiriça em relação à duração normalizada da unidade fronteiriça	0.001 426
20º	emphl	intensidade na janela anterior à fronteira	0.000 584

Tabela 7: Composição do modelo destinado à identificação das fronteiras terminais.

Abaixo, apresentamos a composição dos modelos de identificação de fronteiras não-terminais nas tabelas 8, 9 e 10.

Rank	Sigla	Fenômeno	Peso
1º	zlo	duração da unidade fronteiriça	24.292 13
2º	zr1	duração da unidade pós-fronteiriça	24.187 51
3º	zdloc	mudança na duração da unidade pós-fronteiriça em relação à unidade fronteiriça	24.046 66
4º	psp	presença de pausa	4.013 554
5º	psdur	duração da pausa	2.842 826
6º	ard	mudança na taxa de articulação	0.167 689
7º	sdfod	mudança na variação de F_0 – medida a partir da média de F_0	0.153 763
8º	zl1o	duração da unidade pré-fronteiriça	0.137 096
9º	srd	mudança na taxa de elocução	0.092 402

Tabela 8: Composição do modelo 1 destinado à identificação das fronteiras não-terminais.

Rank	Sigla	Fenômeno	Peso
1º	srl	taxa de elocução na janela anterior à fronteira	0.717 276
2º	sddfol	regularidade do contorno entonacional de F_0 na janela anterior à fronteira – medida através da primeira derivada da mediana de F_0	0.625 759
3º	sdfol	variação de F_0 média na janela anterior à fronteira	0.467 596
4º	ard	mudança na taxa de articulação	0.450 09
5º	fomedl	F_0 média da janela anterior à fronteira	0.374 038
6º	ford	mudança na taxa de proeminência de F_0 (picos de F_0 por segundo)	0.205 596
7º	fomeddloc	reset de F_0	0.098 721
8º	fomedo	F_0 média da unidade fronteiriça	0.087 294
9º	fomedr1	F_0 média da unidade pós-fronteiriça	0.050 374
10º	emphl	intensidade na janela anterior à fronteira	0.012 798

Tabela 9: Composição do modelo 2 destinado à identificação das fronteiras não-terminais.

Rank	Sigla	Fenômeno	Peso
1º	prl	taxa de saliência duracional na janela anterior à fronteira 15	1.555 76
2º	prd	mudança na taxa de saliência duracional 15	0.5872
3º	prr	taxa de saliência duracional na janela posterior à fronteira 14	9.515 03
4º	sdfor	variação de F_0 média na janela posterior à fronteira	0.520 011 6
5º	SDzl	variação no ritmo na janela anterior à fronteira	0.313 216 3
6º	dfomedr1	mudança no contorno entonacional de F_0 na unidade VV imediatamente posterior à fronteira	0.298 542 1
7º	dfomedl10	mudança no contorno entonacional de F_0 na unidade VV imediatamente anterior à fronteira	0.181 148 1
8º	dfomeddloc	mudança no contorno entonacional de F_0 entre as unidades pós-fronteiriça e fronteiriça	0.142 539 6

Tabela 10: Composição do modelo 3 destinado à identificação das fronteiras não-terminais.

Resultados do modelo destinado às fronteiras terminais

Situações de acordo entre segmentadores e modelo TB

Em geral, o modelo destinado à identificação das fronteiras terminais é bastante eficiente. A porcentagem de fronteiras terminais identificadas pelo modelo TB nas duas amostras é bastante compatível com a porcentagem de fronteiras terminais observadas pela maioria dos segmentadores. Os modelos apresentados neste trabalho foram efetivamente construídos com a Amostra I. Com a Amostra I, o modelo identifica 80% das fronteiras prosódicas terminais. Com a Amostra II, observa-se que há uma queda de 6% de desempenho em relação ao desempenho alcançado com a Amostra I. Deste modo, na Amostra II, o modelo identifica 74% das fronteiras terminais.

Este declínio na performance do modelo é provavelmente justificado pelo fato de a amostra originalmente utilizada no desenvolvimento do modelo (Amostra I) possuir uma maior quantidade de fronteiras terminais marcadas com pausa silenciosa, o que não acontece para a Amostra II. Na Amostra I, em um total de 70 posições de TB, 80% é imediatamente seguido por pausa. Na Amostra

II, em um total de 46 posições de TB, apenas 35% é seguido por pausa.

Como já foi dito, a categoria NO-TB engloba NTB e NB. Em relação a esta categoria, os resultados obtidos pelo modelo TB indicam que aproximadamente 66% das posições de NTB e 96% das posições de NB foram corretamente marcadas como NO-TB pelo modelo.

Situações de desacordo entre segmentadores e modelo TB

Os resultados mostram que os desacordos entre segmentadores e modelo TB compreendem duas situações fundamentais. A primeira delas ocorre quando o modelo marca TB ou NTB como NO-TB. Neste caso, observou-se que aproximadamente 21,6% das posições de TB não foram identificadas pelo modelo, pois o modelo marcou incorretamente tais posições como NO-TB. Observou-se que isso ocorre principalmente em posições de TB sem pausa. Além disso, cerca de 30% das posições de NTB marcadas pelos segmentadores foram equivocadamente etiquetadas como TB pelo modelo TB. Neste caso, observou-se que as NTB foram imediatamente seguidas por pausa.

A segunda situação de desacordo ocorre quando o modelo marca NB como TB. Apenas 0,3% das posições de NB foram marcadas como TB pelo modelo. A este respeito, observou-se que as posições de NB são imediatamente seguidas por pausa. De fato, posições seguidas por pausa marcadas como NB comprometeriam toda a análise até o momento desenvolvida, pois é comumente reconhecido que a pausa por si é um parâmetro associado aos diferentes tipos de fronteiras prosódicas.

Por isso, é importante esclarecer que tais posições de NB são na verdade posições de fronteira prosódica. Nestes casos, as posições foram marcadas por quase todos os segmentadores como posições de fronteira prosódica; porém, não há muita clareza entre eles a respeito do tipo de fronteira. Alguns segmentadores marcaram TB, outros NTB, outros *retracting* ou interrupção. Como o BreakDescriptor considera como fronteira prosódica posições em que pelo menos 50% dos segmentadores marcaram necessariamente uma fronteira do mesmo tipo, esse critério não foi completamente atendido. Assim, as posições são consideradas posições de NB, porque os segmentadores se dividiram quanto às categorias anotadas e não alcançaram 50% de acordo entre eles quanto ao tipo de fronteira prosódica percebida. Deste modo, o que parecia ser um grave problema é justificável pela metodologia empregada.[39]

[39] Esse raciocínio é aplicável às posições tratadas pelo *script* como NB que são imediatamente seguidas por pausa e classificadas como NTB pelos modelos NTB 1, NTB 2 ou NTB 3,

Na literatura da área, muitos estudos argumentam que parâmetros relacionados à duração e à F_0 são relevantes para a produção de fronteiras terminais. Neste estudo, os parâmetros relacionados à duração são incluídos no modelo, mas a relevância hierárquica deles dentro do modelo não é grande. Quanto à F_0, o modelo confirma o que é exposto na literatura. No modelo destinado à identificação das fronteiras terminais, os principais parâmetros se relacionam à pausa e à F_0.

De fato, o modelo apresentado parece ser bastante sensível às pausas e tende a etiquetar como TB todas as posições em que ocorre uma pausa, mesmo que a pausa marque NTB. Isso é claramente evidenciado nos casos em que os segmentadores marcaram como NTB e o modelo marcou a posição como TB, porque a NTB em questão era seguida por pausa. A superestimação da pausa também é evidenciada em outra situação. Os resultados mostram que o modelo deixa de reconhecer TB sem pausa e marca tais posições como NO-TB. Assim, NTB com pausa são etiquetadas equivocadamente como TB e TB sem pausa deixam de ser etiquetadas como TB.

Resultados dos modelos destinados às fronteiras não-terminais

Situações de acordo entre segmentadores e modelos NTB

Os modelos destinados à identificação de fronteiras não-terminais apresentados neste trabalho foram efetivamente construídos a partir da Amostra I. Com a Amostra I, o modelo 1 identifica 68% das fronteiras não-terminais, o modelo 2 identifica 25% e o modelo 4 identifica 5%. No total, com o uso dos três modelos, são identificadas, então, 98% das fronteiras não-terminais observadas pelos segmentadores na Amostra I.

Com a aplicação dos modelos 1, 2 e 3 na Amostra II, os resultados obtidos são um pouco diferentes. Com a Amostra II, o modelo 1 identifica 66% das fronteiras não-terminais, o modelo 2 identifica 19% e o modelo 3 identifica 3%. Na Amostra II, os três modelos juntos identificam, então, 88% das fronteiras não-terminais. Deste modo, a performance geral diminui 10% em relação à performance inicial dos modelos com a Amostra I.

Os resultados indicam que o modelo NTB 1, composto principalmente por medidas de pausa e duração normalizada, é o mais ex-

plicativo, porque obteve a maior convergência com a segmentação realizada pelos segmentadores nas duas amostras. Considerando as duas amostras juntas, o modelo NTB 2 explica aproximadamente 22% das fronteiras não-terminais, utilizando principalmente medidas de taxa de elocução e articulação. O modelo NTB 3 explica um número de casos de fronteiras bem menor e suas principais medidas são a taxa de picos de saliência duracional das unidades V-Vs em torno das fronteiras.

O processo de treinamento dos modelos NTB buscou investigar subgrupos de NTB marcadas por diferentes configurações de parâmetros e utilizou as posições de NTB e NB. O procedimento adotado consistiu em desenvolver gradativamente modelos NTB à medida que algumas fronteiras não fossem corretamente identificadas. A reaplicação dos modelos NTB na base completa de dados com todas as posições possíveis (TB, NTB e NB) sem realizar o processo de eliminar fronteiras corretamente identificadas e aplicar o modelo subsequente indica que algumas posições de NTB são reconhecidas pelos três modelos, por dois modelos ou apenas por um modelo NTB. Outras, todavia, não foram identificadas por nenhum dos três modelos NTB. Na Amostra I, a porcentagem de NTB não identificada por nenhum dos modelos corresponde a aproximadamente 2%. Na Amostra II, a porcentagem corresponde a 10%.

O fato de algumas NTB serem identificadas por mais de um dos modelos pode ser explicado em função de que, a partir do momento em que certas fronteiras foram identificadas pelo modelo NTB 1, durante o processo de treinamento dos modelos, eliminou-se a possibilidade dessas fronteiras serem reconhecidas pelos demais modelos, já que elas foram retiradas dos dados. Deste modo, devido à metodologia de desenvolvimento dos modelos empregada, posições de NTB corretamente reconhecidas por NTB 1 sequer tiveram a "oportunidade" de serem reconhecidas pelos modelos NTB 2 e NTB 3, assim como NTB identificadas pelo modelo NTB 2 não tiveram a "oportunidade" de serem reconhecidas pelo modelo NTB 3 subsequente.

Situações de desacordo entre segmentadores e modelos NTB

As situações de desacordo entre segmentadores e modelos NTB são diversas. O primeiro desacordo compreende as situações em que os segmentadores marcam como TB e os modelos marcam como NTB. Observou-se que posições de TB com pausa foram inadequadamente marcadas como NTB por mais de um dos modelos dedicados às NTB. Especificamente, o modelo com maior incidência deste desacordo é o NTB 1. Assim, apesar de o modelo

NTB 1 identificar a maior porcentagem de fronteiras não-terminais nas duas amostras, este modelo marca equivocadamente posições de TB com pausa como NTB. Nos modelos NTB 2 e NTB 3, observou-se que posições de TB com pausa também foram etiquetadas como NTB, em menor proporção, provavelmente, porque estes modelos não são os primeiros aplicados. No caso dos modelos NTB 2 e NTB 3, é importante frisar que estes modelos sequer incluem medidas relacionadas à pausa.

Outro tipo de desacordo diz respeito às posições de TB que são etiquetadas como NB. Neste tipo de desacordo, também há interseções e algumas posições de TB foram marcadas como NB por mais de um dos modelos NTB. Os resultados indicam que este tipo de marcação inadequada é mais frequente nos modelos NTB 2 e NTB 3, porém, isso também ocorre com o modelo NTB 1. No caso do modelo NTB 1, esse tipo de marcação indevida está principalmente relacionado às posições de TB sem pausa. Com a aplicação dos modelos NTB 2 e NTB 3, as posições de TB que são reconhecidas como NB são majoritariamente TB com pausa.

Para as situações em que os modelos marcam (i) TB como NB; (ii) TB como NTB, independente do modelo NTB em questão, é importante discutir se tal desacordo seria realmente um problema, pois todos os três modelos NTB foram desenvolvidos exclusivamente através das posições de NB e NTB. Deste modo, as posições TB não foram levadas em consideração no momento de desenvolvimento dos modelos de identificação de NTB.

A priori, ainda que isso possa comprometer o desempenho de uma possível ferramenta de detecção automática de fronteiras, tal desacordo não deve ser visto como um grave problema, porque qualquer modelo estatístico treinado com a finalidade de identificar determinadas categorias certamente teria uma maior dificuldade na classificação de categorias com as quais ele não foi devidamente treinado. Do ponto de vista da máquina, em relação aos modelos NTB apresentados neste trabalho, presume-se que haverá inerentemente uma maior dificuldade para classificar posições TB.

Para as situações (i) e (ii) descritas acima, naturalmente, a etiqueta atribuída é inadequada e não está de acordo com a etiqueta atribuída pelos segmentadores. Porém, isso afetaria o desempenho da ferramenta de detecção automática de fronteiras de forma moderada, porque a ideia geral deste trabalho consiste em primeiramente aplicar o modelo TB, identificar as posições de TB e retirar do conjunto de dados posições de TB identificadas pelo modelo TB, e, com as posições restantes, aplicar sequencialmente os modelos NTB para então identificar as posições de NTB.

Assim, o total de TB que pode ser marcado equivocadamente como NB ou NTB pelos modelos NTB é limitado à quantidade de

TB que não foi previamente identificada pelo modelo TB. Nesse sentido, em uma possível ferramenta de detecção automática de fronteiras prosódicas construída com base nos modelos apresentados neste estudo, o impacto das situações i e ii não deve ser extremo, pois a performance do modelo TB é alta.

Outro tipo de desacordo diz respeito às posições de NTB que são equivocadamente consideradas posições de NB pelos modelos. Neste caso, certas posições de NTB foram marcadas como NB por mais de um dos modelos NTB. Este tipo de desacordo foi mais recorrente com o uso dos modelos NTB 3 e NTB 2. Com o modelo NTB 1, é natural que a incidência de NTB etiquetada como NB seja menor, pois este é claramente o modelo com melhor performance. A esse respeito, observa-se que a falha do modelo NTB 1 consiste em não reconhecer posições de NTB sem pausa como tal, etiquetando a grande maioria das posições de NTB sem pausa como NB.

Por fim, o último tipo de desacordo entre segmentadores e modelos NTB compreende os casos em que os modelos marcam de forma indevida posições de NB como NTB. De fato, esse tipo de marcação deveria ser visto como um comportamento do modelo totalmente oposto ao comportamento dos segmentadores. Uma análise detalhada dos resultados mostra que algumas posições de NB marcadas erroneamente como TB compreendem posições duvidosas para os segmentadores.

Nestes casos, os segmentadores tiveram dúvidas a respeito da natureza perceptual da fronteira, de modo que a mesma posição foi marcada como TB, NTB, *retracting* e também interrupção. Os segmentadores não alcançaram, então, o acordo de pelo menos 50% quanto ao tipo de fronteira prosódica. Em virtude desta falta de acordo, as posições foram consideradas posições de NB pelo *script* BreakDescriptor para Praat. Do ponto de vista dos modelos, é, então, justificável que os modelos marquem estas posições de NB como NTB, pois as posições em questão seguramente são fronteiras prosódicas. Assim, enquanto os segmentadores não têm acordo suficiente para informar o tipo de fronteira prosódica, os modelos indicam que tais posições são NTB.

Os demais casos de NB que são considerados como NTB realmente são posições de NB. Então, os modelos efetivamente falharam e ainda não há fortes argumentos que possam explicar esse tipo de marcação indevida feita pelos modelos. A este respeito, é importante acrescentar algumas observações.

Em geral, os modelos NTB apresentados identificam a grande maioria das posições de NTB. O principal problema observado reside no fato de que o modelo com a maior capacidade de identificar fronteiras não-terminais (NTB 1) falha ao marcar aproximadamente 1/4 das posições de NB como NTB. De fato, apesar

dos diversos tipos de marcações indevidas citadas ao longo da seção, este é o principal problema dos modelos NTB. Assim, é de suma importância investigar aspectos que possam justificar tais resultados e, ao mesmo tempo, direcionar o que ainda pode ser aprimorado pelo menos no modelo NTB 1 de identificação de NTB.

Perspectivas futuras

O trabalho aqui exposto prevê a realização de uma análise qualitativa mais detalhada a respeito dos parâmetros, além da pausa, que possam explicar as diferentes situações de desacordo entre segmentadores e modelos. De fato, já sabemos que a pausa é superestimada como marca de TB.[40] Os resultados apresentados neste trabalho também sugerem que os modelos desenvolvidos para identificar NTB superestimam a presença de pausa para reconhecer a presença de NTB.

[40] Raso et al., "O papel da pausa na segmentação prosódica de corpora de fala" (2015).

No caso das TB, uma consequência disso, mas não a única, é o menor desempenho do modelo na Amostra II, onde há significativamente menos pausas do que na Amostra I. Além disso, o modelo destinado às TB tende a marcar NTB com pausa como TB e, muitas vezes, não reconhece TB sem pausa como tal. Para as NTB, os modelos também enfrentam dificuldades para lidar com as pausas. No caso dos modelos voltados para as NTB, observa-se que posições de TB com pausa são etiquetadas como NTB e posições de NTB sem pausa frequentemente não são reconhecidas.

Com o intuito de compreender melhor a percepção humana no que diz respeito às fronteiras prosódicas, serão verificados perceptualmente por um lado os desacordos nas posições de maior dúvida para os segmentadores quanto à percepção das fronteiras, e por outro os desacordos (entre segmentadores e modelo) nos casos em que os segmentadores têm acordo alto ou total entre eles. No primeiro caso, a análise, possivelmente, indicará configurações de parâmetros fonético-acústicos que geram uma maior dificuldade no sistema de percepção das fronteiras. Esses casos não devem ser considerados como verdadeiros desacordos entre segmentadores e modelo, pois o desacordo está já presente entre os humanos.

Paralelamente às análises anteriores citadas, será feita uma análise qualitativa perceptual sobre as configurações de parâmetros que podem ter levado os segmentadores e os modelos a atribuírem uma etiqueta do mesmo tipo em uma posição qualquer. Deste modo, para as situações de acordo entre a segmentação do grupo e a segmentação prevista pelo modelo, será verificado perceptualmente se as fronteiras e as demais posições identifi-

cadas pelos modelos são realizadas por certas configurações de parâmetros, que potencialmente favorecem a predição automática dos modelos. Será verificado também se fronteiras não-terminais identificadas por mais de um modelo NTB são aquelas marcadas por um maior número de segmentadores e se são mais salientes do ponto de vista da percepção humana.

Outra etapa deste trabalho consiste em realizar, a partir de dados completamente balanceados, o treinamento de modelos que visam identificar fronteiras prosódicas e suas subcategorias. Em relação às categorias principais (TB e NTB), será realizado o treinamento de modelos a partir de dados completamente balanceados. De fato, normalmente se obtêm resultados melhores quando os modelos são frutos de um treinamento com dados balanceados. Com isso, esperamos melhorar o desempenho principalmente do modelo TB, inclusive reduzindo o número de parâmetros necessários.

Em relação às subcategorias, será realizado o treinamento de modelos que visam identificar subcategorias de TB e NTB, depois de se ter pedido a um número restrito de segmentadores com maior acordo entre eles que marquem (sem modificar a segmentação já feita) também categorias mais finas, sempre com base na percepção. Certamente será pedido que seja marcada a percepção de pausa em caso de TB e NTB. Isso permitirá a elaboração de um modelo para TB com pausa, um modelo para TB sem pausa, um modelo para NTB com pausa e outro para NTB sem pausa, com o objetivo de superar a tendência de superestimar a pausa na determinação das fronteiras prosódicas.

Para tornar mais efetivo o reconhecimento das fronteiras NTB, uma estratégia a ser tentada consiste em separar parte dessas fronteiras com base em características acústicas particularmente salientes à percepção, como a presença de um sinal de continuidade ou a falta de alongamento pré-fronteiriço; outras características podem emergir da análise qualitativa. O objetivo disso é construir sub-modelos que permitam capturar com maior precisão grupos homogêneos de NTB e reduzir assim as fronteiras que precisam ser capturadas com base em modelos mais abrangentes. Esse tipo de estratégia naturalmente implicará a individualização de uma hierarquia entre os diferentes modelos.

Os resultados fornecidos neste estudo mostram, em geral, que os parâmetros relevantes hierarquicamente para a identificação automática das fronteiras prosódicas tendem a estar localizados próximos às próprias fronteiras. Portanto, uma outra estratégia que será implementada no futuro é a redução gradativa da janela de busca. Desta maneira espera-se reduzir o ruído gerado nos modelos. De fato, a extensão da janela estabelecida causa um efeito direto nas medidas globais extraídas pelo *script*. Até o momento,

neste trabalho, a janela de análise utilizada é composta pela própria unidade em análise, 10 unidades V-V anteriores e posteriores em relação à posição analisada. Futuramente, as janelas de trabalho do *script* serão reduzidas com o intuito de avaliar o efeito gerado por essa redução nos parâmetros globais que estão incluídos nos modelos. Assim, a definição da extensão das janelas pode ser vista como uma etapa crucial deste trabalho, pois ela pode auxiliar no desenvolvimento de uma ferramenta de identificação automática de fronteiras cuja implementação seja mais simples e o funcionamento geral seja mais rápido, em termos de processamento computacional.

Agradecimentos

Agradecemos à equipe do Laboratório de Estudos Empíricos e Experimentais da Linguagem da Universidade Federal de Minas Gerais pelo trabalho de percepção e anotação das fronteiras prosódicas. Agradecemos também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) pelos financiamentos que tornaram possível a pesquisa.

Referências

Ananthakrishnan, Sankaranarayanan e Shrikanth S. Narayanan (2005). "An Automatic prosody recognizer using a coupled multi-stream acoustic model and a syntactic-prosodic language model". *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing* (18 de março de 2005–23 de março de 2005). Volume 1. 5 volumes. The Institute of Electrical and Electronics Engineers – Signal Processing Society. Philadelphia, pp. 269–272. ISBN: 0780388747. DOI: 10.1109/ICASSP.2005.1415102.

Auran, Cyril, Caroline Bouzon e Daniel Hirst (2004). "The Aix-MARSEC project: an evolutive database of spoken British English". *Proceedings of Speech Prosody 2004*. Speech Prosody 2004 (23 de março de 2004). Editado por Bernard Bel e Isabelle Marlien. International Speech Communication Association. Nara, pp. 561–564. ISBN: 9782951823310.

Barbosa, Plínio Almeida (1994). "Caractérisation et génération automatique de la structuration rythmique du français". Tese de doutoramento. Grenoble: Institut national Polytechnique.

Barbosa, Plínio Almeida (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes. ISBN: 978-8571132337.

Barbosa, Plínio Almeida (2010). "Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech". *Proceedings of the Fifth International Conference on Speech Prosody*. Speech Prosody 2010 (11 de maio de 2010–14 de maio de 2010). Chicago. URL: https://www.isca-speech.org/archive/sp2010/papers/sp10_067.pdf.

Barth-Weingarten, Dagmar (2016). *Intonation Units Revisited. Cesuras in talk-in-interaction*. Amsterdam: John Benjamins Publishing Company. ISBN: 9789027226396. DOI: 10.1075/slsi.29.

du Bois, John W., Susanna Cumming, Stephan Schuetze-Coburn e Danae Paolino (1992). "Discourse transcription". In: *Santa Barbara Papers in Linguistics*. Volume 4. Santa Barbara: University of California, Santa Barbara. URL: https://www.linguistics.ucsb.edu/research/santa-barbara-papers.

du Bois, John W. et al. (2000). *Santa Barbara corpus of spoken American English*. URL: https://www.linguistics.ucsb.edu/research/santa-barbara-corpus.

Bybee, Joan (2010). *Language, usage and cognition*. Cambridge (UK): Cambridge University Press. ISBN: 9780521851404.

Chafe, Wallace L. (1980). "The deployment of consciousness in the production of a narrative". In: *The pear stories. Cognitive, cultural, and linguistic aspects of narrative production*. Editado por Wallace L. Chafe. Norwood (NJ): Ablex, pp. 9–50. ISBN: 9780893910327.

Chafe, Wallace L. (1994). *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press. ISBN: 0226100545.

Cheng, Winnie, Christopher Greaves e Martin Warren (2005). "The creation of a prosodically transcribed intercultural corpus. The Hong Kong Corpus of Spoken English (prosodic)". *ICAME journal* 29, pp. 47–68.

Cooper, William E. e Jeanne Paccia-Cooper (1980). *Syntax and speech*. Cambridge (MA): Harvard University Press. ISBN: 0674860756.

Cresti, Emanuela (2000). *Corpus di italiano parlato*. Volume 1. Firenze: Accademia della Crusca. ISBN: 9788878850017.

Cresti, Emanuela e Massimo Moneglia (2005). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Volume 15. Amsterdam: John Benjamins Publishing. ISBN: 902722286X.

Croft, William (1995). "Intonation units and grammatical structure". *Linguistics* 33.5, pp. 839–882. DOI: 10.1515/ling.1995.33.5.839.

Cruttenden, Alan (1997). *Intonation*. 2ª edição. Cambridge (UK): Cambridge University Press. ISBN: 0521591821.

Frazier, Lyn, Katy Carlson e Charles Clifton Jr (2006). "Prosodic phrasing is central to language comprehension". *Trends in cognitive sciences* 10.6, pp. 244–249. DOI: 10.1016/j.tics.2006.04.002.

Halliday, M. A. K. (1965). "Speech and situation". *English in Education* 2.A2, pp. 14–17. DOI: 10.1111/j.1754-8845.1965.tb01331.x.

Izre'el, Shlomo, Heliana Mello, Alessandro Panunzi e Tommaso Raso, editores (2020). *In Search of Basic Units of Spoken Language. A corpus-driven approach*. John Benjamins. ISBN: 9789027204974. DOI: 10.1075/scl.94.

Kjelgaard, Margaret M. e Shari R. Speer (1999). "Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity". *Journal of Memory and Language* 40.2, pp. 153–194. DOI: 10.1006/jmla.1998.2620.

Kraljic, Tanya e Susan E Brennan (2005). "Prosodic disambiguation of syntactic structure. For the speaker or for the addressee?" *Cognitive psychology* 50.2, pp. 194–231. DOI: 10.1016/j.cogpsych.2004.08.002.

Ladd, Robert (2008 [1996]). *Intonational Phonology*. 2nd, revised edition. Cambridge: Cambridge University Press. ISBN: 9781139473996.

Mertens, Piet e Anne Catherine Simon (2013). "Towards Automatic detection of prosodic boundaries in spoken French". *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)* (11 de setembro de 2013–13 de setembro de 2013). Editado por Piet Mertens e Anne Catherine Simon. Leuven, pp. 81–87. ISBN: 9789090278766.

Mettouchi, Amina e Christian Chanard (2010). "From fieldwork to annotated corpora. The CorpAfroAs Project". *Faits de Langues* 35–36.2, pp. 255–265. DOI: 10.1163/19589514-035-036-02-900000011.

Mo, Yoonsook, Jennifer Cole e Eun-Kyung Lee (2008). "Naïve listeners' prominence and boundary perception". *Proceedings of the Fourth Conference on Speech Prosody* (6 de maio de 2008–9 de maio de

2008). Editado por Plínio Almeida Barbosa, Sandra Madureira e César Reis. International Speech Communication Association. Campinas, pp. 735–738.

Moneglia, Massimo e Emanuela Cresti (1997). "L' intonazione e i criteri di trascrizione del parlato adulto e infantile". In: *Il Progetto CHILDES-Italia: Contributi di ricerca sulla lingua italiana*. Editado por Umberta Bortolini e Elena Pizzuto. Pisa: Edizioni del Cerro, pp. 57–90. ISBN: 9788882160111.

Ni, Chong-Jia, Ai-Ying Zhang, Wen-Ju Liu e Bo Xu (2012). "Automatic prosodic break detection and feature analysis". *Journal of Computer Science and Technology* 27.6, pp. 1184–1196. DOI: 10.1007/s11390-012-1295-z.

Ostendorf, Mari, Patti J Price e Stefanie Shattuck-Hufnagel (1995). *The Boston University radio news corpus*. URL: https://catalog.ldc.upenn.edu/LDC96S36.

Pierrehumbert, Janet (1980). "The phonetics and phonology of English intonation". Tese de doutoramento. Cambridge (MA): Massachusetts Institute of Technology.

Pierrehumbert, Janet B., Mary E. Beckman e D. R. Ladd (2000). "Conceptual foundations of phonology as a laboratory science". In: *Phonological knowledge. Conceptual and empirical issues*. Editado por Noel Burton-Roberts, Philip Carr e Gerard Docherty. Oxford: Oxford University Press, pp. 273–304. ISBN: 9780199245772.

de Pijper, Jan Roelof e Angelien A. Sanderman (1994). "On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues". *The Journal of the Acoustical Society of America* 96.4, pp. 2037–2047. DOI: 10.1121/1.410145.

Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Raso, Tommaso e Heliana Mello (2012). *C-ORAL-BRASIL I: corpus de referência do Português Brasileiro falado informal*. Belo Horizonte: Universidade Federal de Minas Gerais. ISBN: 9788570419439.

Raso, Tommaso, Heliana Mello e Lúcia Ferrari (sem data). *C-ORAL-BRASIL II: corpus de referência do Português Brasileiro falado informal*. Em preparação.

Raso, Tommaso, Maryualê Malvessi Mittmann e Anna Carolina Oliveira Mendes (2015). "O papel da pausa na segmentação prosódica de corpora de fala". *Revista de Estudos da Linguagem* 23.3, pp. 883–922. DOI: 10.17851/2237-2083.23.3.883-922.

Reed, Beatrice Szczepek (2012). "Prosody, syntax and action formation. Intonation phrases as action components". In: *Prosody and embodiment in interactional grammar*. Editado por Pia Bergmann, Jana Brenning, Martin Pfeiffer e Elisabeth Reber. Berlin: Walter de Gruyter, pp. 142–169. ISBN: 9783110295047.

Ross, Kenneth N. e Mari Ostendorf (1996). "Prediction of abstract prosodic labels for speech synthesis". *Computer Speech & Language* 10.3, pp. 155–185. DOI: 10.1006/csla.1996.0010.

Schafer, Amy J., Shari R. Speer e Paul Warren (2005). "Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task". In: *Approaches to studying world-situated language use. Bridging the language-as-product and language-as-action traditions*. Editado por John C. Trueswell e Michael K. Tanenhaus. MIT Press, pp. 209–225. ISBN: 9780262201490.

Schubiger, Maria (1958). *English intonation. Its form and function*. Halle: M. Niemeyer Verlag.

Schuetze-Coburn, Stephan Mark (1994). "Prosody, syntax, and discourse pragmatics. Assessing information flow in German conversation". Tese de doutoramento. University of California, Los Angeles.

Schuurman, Ineke, Machteld Schouppe, Heleen Hoekstra e Ton van der Wouden (2003). "CGN, an annotated corpus of spoken Dutch". *Proceedings of 4th International Workshop on Linguistically Interpreted Corpora (LINC-03) at EACL 2003* (13 de abril de 2003–14 de abril de 2003). Editado por Anne Abeillé, Silvia Hansen-Schirra e Hans Uszkoreit. Association for Computer Linguistics. Budapest, pp. 101–108. URL: https://www.aclweb.org/anthology/W03-2414.

Selkirk, Elisabeth (2005). "Comments on intonational phrasing in English". In: *Prosodies. With special reference to Iberian languages*. Editado por Sónia Frota, Marina Vigário e Maria João Freitas. Berlin: Walter de Gruyter, pp. 11–58. ISBN: 9783110184440.

Silverman, Kim et al. (1992). "ToBI: A standard for labeling English prosody". *Proceedings of the 7th International Conference on Spoken Language Processing*. 7th International Conference on Spoken Language Processing (16 de setembro de 2002–20 de setembro de 2002). Editado por John H. L. Hansen e Bryan Pellom. International Speech Communication Association. Denver, pp. 867–870.

Snedeker, Jesse e John Trueswell (2003). "Using prosody to avoid ambiguity. Effects of speaker awareness and referential context". *Journal of Memory and language* 48.1, pp. 103–130. DOI: 10.1016/S0749-596X(02)00519-3.

Speer, Shari R., Margaret M. Kjelgaard e Kathryn M. Dobroth (1996). "The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities". *Journal of psycholinguistic research* 25.2, pp. 249–271. DOI: 10.1007/BF01708573.

Swerts, Marc (1994). "Prosodic Features of Discourse Units". Tese de doutoramento. Technische Universiteit Eindhoven. DOI: 10.6100/IR411593.

Swerts, Marc (1997). "Prosodic features at discourse boundaries of different strength". *The Journal of the Acoustical Society of America* 101.1, pp. 514–521. DOI: 10.1121/1.418114.

Teixeira, Bárbara Helohá Falcão (2018). "Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea". Tese de mestrado. Universidade Federal de Minas Gerais. URL: http://hdl.handle.net/1843/LETR-AX8HUG.

Warren, Paul, Esther Grabe e Francis Nolan (1995). "Prosody, phonology and parsing in closure ambiguities". *Language and cognitive processes* 10.5, pp. 457–486. DOI: 10.1080/01690969508407112.

Watson, Duane e Edward Gibson (2004). "The relationship between intonational phrasing and syntactic structure in language production". *Language and cognitive processes* 19.6, pp. 713–755. DOI: 10.1080/01690960444000070.

Wightman, Colin W. e Mari Ostendorf (1994). "Automatic labeling of prosodic patterns". *IEEE Transactions on speech and audio processing* 2.4, pp. 469–481. DOI: 10.1109/89.326607.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf e Patti J. Price (1992). "Segmental durations in the vicinity of prosodic phrase boundaries". *The Journal of the Acoustical Society of America* 91.3, pp. 1707–1717. DOI: 10.1121/1.402450.