

ESTRATÉGIAS DE FINALIZAÇÕES FRASAIS E A ANÁLISE AUTOMÁTICA DA ENTOAÇÃO: UM ESTUDO PILOTO

STRATEGIES FOR PHRASING AND AUTOMATIC ANALYSIS OF INTONATIONAL: A PILOT STUDY

Waldemar FERREIRA NETTO¹

Marcus Vinícius Moreira MARTINS²

Resumo: O propósito deste artigo é estabelecer critérios e procedimentos para a segmentação e classificação frasais automáticas da língua portuguesa no âmbito do projeto de pesquisa ExProsodia. Baseados em levantamento bibliográfico, desenvolvemos um conjunto de nove parâmetros objetivos e específicos, que se referem a aspectos da marcação de fronteiras de frases entoacionais – direção, alvo, movimento, *shift*, alongamento e pausa – e aspectos do intervalo que se estabelece entre frases entoacionais e a entoação global do discurso – correlação, distância e variação de foco e ênfase. Aplicamos esses parâmetros a uma gravação para análise automática e foi possível verificar que a impossibilidade de se fazer comparações diretas e completamente inequívocas entre intuição humana e análise instrumental automática não permite qualquer tipo de construção subjetiva. Assim, os resultados obtidos por esse tipo de análise, para que sejam interpretados linguisticamente, têm de receber um tratamento adequado às necessidades humanas compartilhadas, por sua vez, por um grupo de falantes da mesma língua.

Palavras-chave: Prosódia. Segmentação. Entoação. Análise Automática.

Abstract: This paper aims to establish criteria and procedures to parse and automatically classify phrases in Portuguese for the research project called ExProsodia. Based on the literature review, we developed a set of nine objective and specific parameters referring to aspects of establishing boundaries of intonational phrases – direction, target, movement, shift, lengthening, and pause – and aspects of the interval between intonational phrases and intonation of the discourse – correlation, distance and focus variation and emphasis. We have applied these parameters in a recording for automatic analysis and one could see that the impossibility to make direct and completely correct comparisons between human intuition and automatic instrumental analysis allows no kind of subjective construction. Thus, to be linguistically read, the results for this kind of analysis have to be treated according to shared human needs by a group of speakers of the same language.

Keywords: Prosody. Phrasing. Intonation. Automatic Analysis.

¹ Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; wafnetto@usp.br; <http://orcid.org/0000-0002-4136-341X>

² Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; marcusvmmartins@gmail.com; <http://orcid.org/0000-0002-6835-5639>

Introdução

O propósito deste artigo é estabelecer critérios e procedimentos para a segmentação e classificação automática do que chamamos de frases entoacionais, no âmbito do projeto de pesquisa ExProsodia. Nosso objetivo principal é analisar as estratégias de finalização utilizadas pelos falantes, a partir de parâmetros acústicos adotados em Martins e Ferreira Netto (2018). É importante destacar que entendemos que estas frases entoacionais não coincidem necessariamente com as fronteiras de sintagmas da língua portuguesa já descritos. Nossa intenção não é estabelecer um modelo para a análise de *phrasing*. Ademais, entendemos que os parâmetros acústicos, além de estabelecerem pontos de segmentação, trazem indícios para uma possível classificação destas frases segundo intenções específicas ou estados do falante.

A relação entre prosódia e o estabelecimento de fronteiras é amplamente discutida na literatura. No entanto, a relação entre a entoação, a finalização e a estruturação das frases entoacionais ainda é incipiente e, só mais recentemente, algumas investigações científicas têm trazido novidades para a definição dessa entoação na língua portuguesa (OLIVEIRA, 2002; OLIVEIRA JR., 2003; FERREIRA NETTO; CONSONI, 2008; FERREIRA NETTO; CONSONI, PERES, 2009; BAZ, 2011; BARBOSA, 2013; ROSA, 2015; MITTMANN; BARBOSA, 2016; ROSA; CONSONI; FERREIRA NETTO *et al.*, 2016; SILVA, 2017; TEIXEIRA; BARBOSA; RASO, 2018). De maneira geral, os critérios estabelecidos por estes trabalhos tratam de variações tonais em posição final de frases, seguidas ou não de pausas, que caracterizariam fronteiras de frases entoacionais.

Baseados na discussão bibliográfica que virá a seguir, desenvolvemos um conjunto de nove parâmetros para serem aplicados na análise e na interpretação da entoação frasal e discursiva global da língua portuguesa. Esses parâmetros referem-se a aspectos da marcação de fronteiras de frases entoacionais – direção, alvo, movimento, *shift*, alongamento e pausa – e aspectos do intervalo que se estabelece entre frases entoacionais e a entoação global do discurso – correlação, distância e variação de foco e ênfase. Para a aplicação efetiva desses parâmetros, aprimoramos o aplicativo também chamado ExProsodia.

Este artigo se organiza de maneira a apresentar, primeiro, as discussões teóricas de que nos servimos para a definição dos parâmetros de segmentação e de classificação de frases entoacionais; segundo, da descrição detalhada dos parâmetros que foram definidos; e, terceiro, da aplicação desses parâmetros em dados diversos. Esta terceira parte do artigo tem como propósito exemplificar as possibilidades que a aplicação do modelo pode oferecer à análise automática da entoação na língua portuguesa, bem

como chamar a atenção para seus limites e necessidade de ajustes. Finalmente, faremos algumas considerações acerca dessas possibilidades, levando em conta as dificuldades que poderão surgir quando de sua aplicação em dados concretos de diferentes línguas, dialetos ou registros. Também faremos sugestões para desenvolvimentos futuros tanto do modelo quanto de sua aplicação.

Parâmetros de segmentação

Nas subseções a seguir, apresentamos os 9 parâmetros utilizados para a segmentação das frases entoacionais, sendo eles: foco/ênfase, alvo, direção, movimento, *shift* entoacional, pausa, alongamento, correlação e diferença. Tais parâmetros são obtidos a partir da extração do f_0 e da análise decomposicional feita utilizando a função (1) para o Tom Médio (TM) (FERREIRA NETTO, 2016):

$$TM_{(t)} = \frac{(t - 1) * TM_{(t-1)} + Z_{(t)}}{t} \quad (1)$$

Na equação (1), t corresponde a cada um dos momentos de f_0 em que se faça a medição das frequências e $Z_{(t)}$ é cada um dos valores obtidos nesse momento t . A seleção desses momentos t faz-se pelas restrições adequadas a cada língua quanto à duração mínima, à intensidade mínima e ao intervalo de frequência. O fator $(t - 1)$ estabelece o número total dos momentos anteriores em relação ao atual como peso na média geral de modo a calcular a média acumulada no tempo. A cada uma das frequências Z obtidas nos momentos t convencionou-se chamar de UBI (*unidade básica de entoação*). Deste modo, o Tom Médio (TM) é definido como a série temporal de UBIs que se obteve a partir da decomposição apresentada em (1).

O tom médio é o parâmetro básico para o estabelecimento dos limites laterais superior e inferior, definidos a partir do conceito de limiares de diferenciação tonal (l.d.t.). De acordo t'Hart e *et al.* (1990), variações tonais acima das l.d.t.'s são perceptivamente salientes para os falantes e concluem que para o holandês é necessária uma alteração da ordem de 3 semitons para que esse efeito ocorra. Consoni (2011) e Martins e Ferreira Netto (2017) encontraram valores semelhantes para o Português Brasileiro, de modo que os limites laterais são definidos pelas equações (2) e (3):

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

$$L_{Sup} = 1,19 * TM_{(t-1)} \quad (2)$$

$$L_{Inf} = 0,84 * TM_{(t-1)} \quad (3)$$

As UBIs que extrapolam o limite superior se caracterizam pela maior saliência tonal e, portanto, estabelecem o que se convencionou chamar de *Foco/Ênfase superior (F/Esup)*. Por sua vez, as UBIs cujos valores ocorram abaixo do limite inferior da faixa de frequência do Tom Médio caracterizam-se pela atenuação da sonoridade da série tonal (KOHLEH; NIEBUHR, 2007); e estabelecem o *Foco/Ênfase inferior (F/Einf)*. Por fim, a alternância de produção de UBIs dentro desta faixa definida pelos limites laterais é chamada de Ritmo Tonal, de modo que a interrupção do ritmo tonal é tomada como uma pausa intencional, a depender da duração da interrupção. Abaixo, na tabela 1, estão arrolados os valores utilizados para se considerar cada parâmetro em nossa análise.

Tabela 1. Parâmetros utilizados para análise automática dos parâmetros: direção, movimento e alvo

Direção	Descendente $\frac{f0_{(t-1)}}{f0_{(t)}} > 1$	Ascendente $\frac{f0_{(t-1)}}{f0_{(t)}} < 1$
Movimento	Oblíquo $\frac{f0_{(t-1)}}{f0_{(t)}} > 0,84$ e $\frac{f0_{(t-1)}}{f0_{(t)}} < 1,19$	Por salto $\frac{f0_{(t-1)}}{f0_{(t)}} < 0,84$ ou $\frac{f0_{(t-1)}}{f0_{(t)}} > 1,19$
Alvos	Autêntico $\frac{f0_t}{TM_t} > 0,84$	Plagal $\frac{f0_{(t)}}{TM_{(t)}} < 0,84$

Fonte: Elaboração própria

Finalização: alvo, direção, movimento

Estabelecido o tom médio e seus intervalos laterais, temos um terceiro componente a ser considerado, o qual se relaciona com as fronteiras de uma frase entoacional, convencionado como Finalização (F). A partir dele podemos estabelecer outros parâmetros importantes para a nossa análise, a saber: *direção, alvo e movimento* da entoação. Em nosso entender, esses valores se relacionam diretamente com a variação final de f_0 em relação a um valor médio estabelecido para as frases entoacionais, no caso o tom

médio. Dessa maneira, entende-se que a direção de f_0 , com relação à finalização, pode ser ascendente ou descendente. Jones (1922), Troubetzkoy (1973 [1939]) e Pike (1945) eram concordes quanto ao fato de que a variação entoacional no correr da frase, bem como em sua finalização, ascendente ou descendente, estabeleciam sentidos suplementares e poderiam atuar com função conclusiva ou de continuidade, respectivamente. A proposta de Pike (1945), que estabelecia um conjunto de 4 tons para a língua inglesa, previa uma diferença na finalização de frases entoacionais, quanto ao alvo da entoação descendente. Tons mais graves, de número 4, seriam prontamente relacionados a uma finalização, do que tons menos graves, de número 3, que se associariam a uma noção de continuidade ou de hesitação. Acima deste, tons de número 1 ou 2, mais agudos, estariam associados ou a locuções interrogativas ou à incompletude da locução. Di Cristo (1981) chamou a atenção para o fato de que, em sua pesquisa sobre o francês, a finalização sentencial afirmativa ocorria com um intervalo de quarta descendente quando se tratava de fronteira prosódica terminal.

Sobre o *alvo* e *movimento*, propomos uma relação interdisciplinar com a abordagem musical (FERREIRA NETTO, 2017a), assim assumimos que a terminologia musicológica é adequada para caracterizar esse aspecto da entoação em final de frase entoacional. Schoenberg (2001b) faz a diferença entre uma cadência harmônica plagal e uma cadência autêntica. Segundo ele, uma cadência plagal conclui frases musicais “com menor clareza e poder de confirmação do que a cadência autêntica” (SCHOENBERG, 2011b, p. 428). Desse ponto de vista, podemos entender a noção de *alvo* como “autêntico” se a finalização da frase entoacional apresentar-se mais grave do que o limite inferior do tom médio. A variação em final de frase que não se mostrar como tal será considerada “plagal”.

O movimento final de f_0 pode ser oblíquo ou por salto. Schoenberg (2001a, p. 27) define movimento oblíquo como “aquele em que uma das vozes se move enquanto a outra fica parada.”, isto é, no caso de um movimento oblíquo, os valores finais de f_0 serão próximos entre si; no caso de um movimento por salto, os valores finais de f_0 serão distantes entre si, ou seja, ocorrerão por salto entre UBIs contíguas. Portanto, as UBIs finais, ou seja, os *alvos* autêntico ou plagal, estarão em diferentes alturas; para que ocorra a finalização é necessário um *movimento* por salto ou oblíquo, cuja *direção* pode ser ascendente ou descendente.

No caso do alvo autêntico, há um movimento entoacional descendente que atinge um ponto grave, abaixo do limite lateral inferior de f_0 , definido pela equação (2). Por sua vez, o alvo plagal caracteriza-se pela curva entoacional dentro dos limites da faixa tonal estabelecida pelo tom médio, ou mesmo acima disso. Nesse caso, frases interrogativas

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

caracterizam-se como finalizações plagais com marcação de foco/ênfase. Entendemos que a finalização por movimento descendente, por salto e autêntico seja a forma mais marcada da finalização de frases assertivas.

Shift

Outro parâmetro, por nós considerado, relacionado à porção final de uma frase entoacional é o *shift entoacional*. Fairbanks (1940, p. 458) verificou que a variação de tons em *shift*, ou seja, a mudança tonal que ocorre entre o tom final de uma fonação e o inicial da subsequente, na leitura “normal”, era de 1,8 tons. Na investigação de Snidecor (1943), o autor verificou que a variação de tons em *shift* ocorria com intervalos ascendentes de 2,03 tons e descendentes de 1,84 tons, na fala improvisada; e intervalos ascendentes de 2,15 tons e descendentes de 1,68 tons, na leitura. De maneira geral, houve uma pequena variação invertida entre leitura e fala improvisada quanto ao *shift* que, entretanto, não atingiu 1 semitom. Em ambas as pesquisas, os valores eram tomados de f_0 a cada 38 milissegundos. Em trabalho posterior, Snidecor (1944) verificou que as diferenças nos valores de *shift* eram maiores quando ocorriam entre frases do que quando ocorriam internamente às frases. Essa diferença teve um valor especialmente grande na delimitação frasal da leitura, atingindo uma média maior do que 9 semitons descendentes, em contraste com um valor médio oscilando entre 5 e 6 semitons para os demais casos, ascendentes ou descendentes. Com base nesses dados, o autor concluiu haver o predomínio da variação descendente de tons interfrasais.

Esses resultados, tal como ele próprio apontou, foram na direção contrária de estudos anteriores aos dele como o de Pronovost (1939), em que as diferenças de valores ascendentes ocorriam em 74% de *shift*; de acordo com o Snidecor (1944), essa diferença decorreria da natureza do material utilizado. Em pesquisa subsequente, Snidecor (1951) encontrou a mesma distribuição de *shift* em leituras feitas por sujeitos selecionados de gênero feminino, corroborando seu trabalho anterior. Nessa pesquisa com a fala feminina comparou-a com a fala masculina e verificou que a fala feminina teve uma variação de *shift* ascendente no interior das sentenças menor do que a masculina. O critério usado pelo autor para a definição de frases considerou como delimitadores as pausas interfrasais com duração igual ou maior a 0,5 s.

No caso desta pesquisa, entendemos que a retomada do tom médio, que se caracteriza pela presença de *shift* ou de *reset*, é estabelecida pela diferença entre a frequência final de uma frase entoacional e a frequência inicial da frase seguinte. O valor dessa diferença, por sua vez, pressupõe que o locutor tenha abandonado o tom médio

da frase entoacional anterior, ou porque já a tinha finalizado, ou por conta de alguma interrupção do ritmo tonal, e promoveu seu reinício. Essa retomada pode ser definida na equação (1) pelo reinício do momento t do TM no multiplicador $(t - 1)$ do TM , bem como no do denominador t . Na medida em que a retomada interrompe o fluxo entoacional da fala, entende-se que, havendo uma variação maior do que a que se define pela equação (2), trata-se de uma forma mais marcada do que sua incoerência em frases assertivas.

Pausa

A noção de pausa como delimitador interfrasal já havia sido referida por Hockett (1953) quando afirmou que uma sentença, por definição, inicia e termina com uma pausa. Algum tempo depois, Snidecor (1944) comparou dados extraídos de percepção e de produção de fala improvisada e de leitura das transcrições dessas mesmas falas. Dentre os resultados, observou que a maioria dos sujeitos (50%) reconhecia de maneira uniforme durante as audições pausas com duração maior do que 150 milissegundos. Nos dados que apresentou, somente pausas com mais de 500 milissegundos foram reconhecidas de maneira uniforme nas gravações por 95% de seus sujeitos. Essa mesma duração de pausa ainda foi utilizada nas primeiras tentativas de segmentação frasal em análises automáticas da entoação feitas por Verzeano e Finesinger (1949) e, posteriormente, por Hargreaves e Starkweather (1959). Ainda nessa mesma linha, Boomer e Dittmann (1962), baseados em testes de percepção, verificaram que pausas de hesitação eram percebidas adequadamente a partir de 500 ms. Junturas delimitadoras interfrasais, entretanto, foram percebidas entre 200 e 500 ms, com discriminação muito menor do que as pausas de hesitação.

A correlação entre variações de frequência e pausas como delimitadores interfrasais que Snidecor (1944) esboçou só seria retomada no trabalho de Trager e Smith (1957), que seguiu a hipótese de Pike (1945) que propunha explicar a entoação do inglês a partir de quatro tons pontuais subjetivos. Trager e Smith (1957) estabeleceram duas noções que seriam usadas como referência posteriormente: a juntura terminal e a frase fonêmica. A juntura terminal se caracterizaria como um movimento de *pitch* ascendente, descendente ou sem variação, nos mesmos termos propostos por Jones (1922) ou por Trubetzkoy (1973 [1939]), mas precedendo um silêncio terminal (TRAGER; SMITH, 1957). A frase fonêmica se caracterizaria por um enunciado mínimo portador de uma única juntura terminal (TRAGER; SMITH, 1957). Garvin e Mathiot (1958) fizeram proposta semelhante na qual, em termos da associação de vários intervalos de tom com a pausa mais longa, três tipos de juntura terminal (parágrafo final, hesitação e junção de transição) poderiam ser estabelecidos contra um único tipo de juntura medial. Daneš (1960), nesse mesmo período, salientou

que, em suas funções de delimitação e de segmentação, os padrões entoacionais não operam independentemente, mas formam, juntamente com a pausa, um dispositivo linguístico especial – chamado juntura. Dittmann e Llewellyn (1967) verificaram que os ouvintes interagem com seus locutores por meio de expressões breves, principalmente após frases fonêmicas completas seguidas de pausas de juntura terminal; os resultados reportados pelos autores corroboram a interpretação de que, além da sintaxe, a marca prosódica também estaria envolvida no reconhecimento da frase fonêmica pelo ouvinte.

Bolinger (1989) chamou a atenção para o fato de que variações entoacionais, taxa de elocução, pausas e a relação tonal entre início e final de sentenças são fatos prosódicos que, em seu conjunto, servem como pistas para a demarcação de unidades informacionais no discurso. Segundo ele, seria um caráter universal da prosódia. Pierrehumbert (1983), baseada em seu próprio modelo de análise entoacional do inglês (PIERREHUMBERT, 1980), procurou esboçar um esquema de análise automática da entoação em que as variações tonais são categorizadas por apenas dois tons – H(igh) e L(ow), além de um terceiro, intermediário, não marcado, definidos por inspeção visual de curva de f_0 – e, dentre outras possibilidades, pela posição final como um delimitador sentencial, marcado como %. A autora considera a declinação como parte integrante da entoação.

Taylor (1994) procurou estabelecer um modelo de análise prosódica automática baseando-se especialmente na variação tonal. Para tanto, desenvolveu uma rotina de trabalho que analisava porções de 50 ms de f_0 – suavizado (*smoothed*) principalmente quanto às interrupções causadas pela ocorrência de segmentos não vozeados – considerando as variações tonais ascendente, descendente ou nula. Taylor (2000), ao desenvolver um modelo de análise automática da entoação, propôs que eventos entoacionais se caracterizassem na cadeia entoacional [*'intonational stream'*] como *pitch accents*, *boundary tones* e uma sequência contígua de ambos: *pitch accents* seriam os deslocamentos de f_0 usados pelo falante para dar algum grau de ênfase a alguma sílaba ou palavra e *boundary tones* também seriam deslocamentos de f_0 , mas para marcar fronteiras de sintagmas, bem como para marcar efeitos de afirmação, continuação ou de questionamento. Segundo ele, eventos entoacionais não poderiam se interromper e, portanto, teriam de ser tratados como uma unidade. Embora propusesse que eventos entoacionais fossem unidades alinhadas com sílabas ou palavras, o autor propôs que, diferentemente da cadeia segmental, a cadeia entoacional fosse um fenômeno contínuo, mesmo na sua relação com os possíveis significados que poderia apresentar. Saliente-se que, nesse modelo de análise automática, intervalos entre eventos entoacionais foram considerados somente como referências para os deslocamentos de f_0 .

A discussão de Taylor (2000) quanto à impossibilidade de se tratar as variações de f_0 de uma argumentação convincente até então parece não justificar a crítica que faz às pesquisas que demonstraram que a entoação poderia ser tratada por padrões pontuais fixos – ou até pelo modelo de *perceptual magnet effect* (KUHL, 1991), como o fez para a cadeia lexical. As fronteiras semânticas associadas a padrões entoacionais foram descritas com bastante precisão no português (CAGLIARI, 1981; MORAES, 2007; MORAES; RILLIARD, 2018). Ao que parece, a diferença entre as cadeias lexical (segmental) e prosódica não poderia se resolver somente pelo aspecto da arbitrariedade e da continuidade, ainda que se possa encontrar essa diferença entre elas. Pode-se entender que a cadeia prosódica é contínua porque tem uma fonte contínua de alimentação (FERREIRA NETTO *et al.*, 2017). Suas unidades podem ser padronizadas por diferentes variações de qualidade e, assim, serem tratadas descontinuamente. No entanto, a cadeia prosódica, seja no que diz respeito à entoação ou ao ritmo, pode ter um uso contínuo, em que os “significados” de suas unidades sonoras variam concomitantemente – isso é bem claro na entoação emocional (mais medo-mais agudo, maior taxa de enunciação; menos medo-menos agudo, menor taxa de enunciação) (OHALA, 1984; SCHERER, 1995; BANSE; SCHERER, 1996) – mas não em expressões padronizadas (e talvez por isso mesmo) como a manifestação de dúvida, de surpresa, etc.

Como citamos nas seções anteriores, no âmbito do modelo ExProsodia, a noção de pausa é interpretada como uma interrupção do ritmo tonal. Esta interrupção pode ser causada tanto pela interferência de um interlocutor, em decorrência de troca ou de tomada de turno, como pela necessidade do próprio locutor. Nesse caso, pode tanto ocorrer a formação de sons de vária natureza, como “pausas preenchidas” ou *fillers*, como a formação de intervalos de silêncio chamados pausas. Uma vez que as pausas que ocorrem na fala podem corresponder a fronteiras de frases entoacionais ou a momentos em que o locutor procura dar alguma expressividade a sua fala, ou ainda a uma tentativa de reorganização de seu discurso, elas não se permitem serem tomadas como marcadores de finalização de frases entoacionais, por si só; mas podem ser tomadas como cruciais para o conjunto de marcadores que atuam como fronteiras de frases entoacionais.

Mais do que marcadores de finalização, pausas também atuam como barreiras para o fluxo da entoação justamente porque interrompem o ritmo tonal. Ainda que ora isso seja especulativo, pode-se assumir que uma finalização de frase entoacional que prescindir de uma pausa exigirá uma compensação quanto às formas marcadas de movimento, direção e alvo tonal, bem como do alongamento de UBI. Para este artigo, entretanto, as pausas serão consideradas obrigatórias na formação de fronteiras de frases entoacionais. Nesse caso, segue-se o critério de que já se tratou anteriormente, entendendo que pausas

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

longas, mais de 500 ms, são mais facilmente percebidas do que as com menor duração; no entanto, as pausas longas ocorrem principalmente em momentos de hesitação. Assim, pressupõe-se que pausas longas sejam mais marcadas como fronteiras do que pausas com menor duração em frases assertivas. Neste artigo não trataremos das “pausas preenchidas”.

Em relação à duração da pausa, pesquisas indicaram que a partir de 200 ms, a pausa é mais saliente (DUEZ, 1985; FLETCHER, 1987; FRIEDMAN; OCONNELL, 1991; OLIVEIRA, 2002; SILVA, 2017). Além disso, Boomer e Dittmann (1962), como já vimos, optaram por diferenciar pausas de hesitação, que ocorriam com maior duração de, pelo menos, 50 ms, e pausas delimitadoras frasais, que se manifestavam já a partir dos 200 ms. Para este artigo, optamos pela reprodução desse princípio, valendo-nos de pausas com duração de 200 ms como marcadores prosódicos liminares de fronteiras interfrasais e pausas como duração maior ou igual a 500 ms como marcadores intencionais de segmentação frasal.

Alongamento

Outro parâmetro considerado na literatura como elemento de marcação de fronteiras é o alongamento de vogal final. Delattre (1966) comparou a duração silábica em posição final e não final acentuada e não acentuada de locuções que formassem grupos de sentido, das línguas alemã, francesa, espanhola e inglesa. Suas análises apontaram uma diferença significativa na duração silábica em posição final e não final, bem como entre sílabas acentuadas e não acentuadas. De maneira geral, em relação às sílabas acentuadas, estas eram mais longas do que as não acentuadas em qualquer posição, mas, em posição final, eram mais longas do que em posição não final. Em relação às sílabas não acentuadas, a situação foi semelhante, apenas diferenciando a língua espanhola que não apresentou essa diferença de forma significativa. A língua francesa, por apresentar acento fixo no final, não permitiu a mesma comparação. Lindblom (1968), analisando a língua sueca, encontrou fenômeno semelhante, mas entendeu que esse fenômeno talvez pudesse relacionar-se mais propriamente à duração das sílabas nas palavras. Oller (1973) chegou às mesmas conclusões em relação às sílabas em final de sentenças inglesas.

Price e seus colegas (1989), com o propósito de estabelecer critérios de segmentação interfrasal automática, selecionaram como delimitadores a duração do fonema, as pausas respiratórias e a variação tonal descendente. Campbell (1993) verificou que a análise dos alongamentos silábico e segmental da língua inglesa poderia ser usada como delimitador prosódico tanto para a segmentação automática entre sintagmas e entre frases. Para tanto, valeu-se de dois critérios: a duração da sílaba como uma só unidade, e a duração dos segmentos internos (*onset*, *peak*, *coda*) da sílaba. Baseado nesses

critérios, verificou que a duração dos segmentos silábicos como uma só unidade variava conforme sua posição sentencial final ou não e que a duração dos segmentos silábicos internos variava em relação à sua posição em fronteira sintagmática.

Embora uma variação de duração segmental, por sua vez, não envolva qualquer relação com a variação entoacional da fala, pode ser tomada como um aspecto interno da formação das UBIs, que, justamente, portam a entoação frasal. Essa variação de duração de UBI final, ora chamada de alongamento, recebeu pouca atenção na prosódia do português falado no Brasil. Assim, esse alongamento se permite avaliar somente em comparação com as demais durações de UBI; caso a UBI final tenha uma duração maior do que a média esperada para as UBIs da mesma frase entoacional em que se encontra, será tratada como forma marcada de finalização de frase assertiva. Para essa avaliação, pode-se utilizar a estatística z .

Indícios classificatórios para a entoação frasal: correlação e diferença

Lieberman (1967) retomou o conceito de *breath-group* que Jones (1922) e Stetson (1927) haviam explorado e pressupôs que, se a tensão dos músculos laríngeos não for deliberadamente aumentada no final da expiração, quando a pressão do ar subglótica diminui, a frequência fundamental da fonação também cairia. Segundo ele, esse decaimento, por sua vez, produziria um padrão prosódico que caracterizaria as fronteiras de sentenças declarativas na fala coloquial. Cohen e t'Hart (1967), analisando o holandês, propuseram que a entoação se realizaria num movimento tonal descendente durante a produção da fala, denominando-o declinação. Segundo os autores, a declinação não estaria necessariamente vinculada a nenhuma unidade linguística como a sentença, mas ao conjunto geral da fala. Para Vaissière (1974), analisando a prosódia do francês, as variações de f_0 teriam essencialmente função demarcadora de sentenças em conjunto com pausas e maior duração dos padrões finais de sílaba f_0 . Maeda (1976) verificou que a língua inglesa também poderia ser interpretada pela declinação, que chamou de *baseline*.

Com proposta semelhante à de Lieberman (1967), o autor considerou que a *baseline* estaria associada à noção de *breath-group* e não propriamente à sentença. Assim, com ligeira diferença em relação ao conceito de declinação de Cohen e t'Hart (1967), Maeda (1976) entendia que os falantes respiravam após cada sentença isolada e, eventualmente, durante a prolação de uma sentença mais longa, e, nessas situações, uma nova *baseline* se desenvolveria; a esse procedimento ele chamou de *reset*. Cooper e Sorensen (1977, 1981) procuraram desvincular esse procedimento da noção de *breath-group*, atribuindo-o a uma opção do próprio locutor, uma vez que, segundo eles, seria acionado diretamente

pela presença do limite sintático e não dependeria de pausa respiratória. Vaissière (1983, p. 86) chamou a atenção para alguns fenômenos prosódicos comuns a várias línguas como, dentre outros, a tendência natural para valores descendentes de f_0 para caracterizar frases declarativas neutras e os movimentos de f_0 correspondentes a um aumento de tensão nas pregas vocais (*reset* da *baseline* e do f_0 local) associarem-se à noção de início e as quedas tonais associarem-se a finalizações. Ladd (1984) fez uma abordagem detalhada de todas essas hipóteses.

A correlação entre tom médio e tom médio total, a diferença entre o tom médio e o tom médio total e a dispersão de entoação acima do tom médio são fatores que, diferentemente dos critérios anteriores usados como fronteiras de frases entoacionais, não se baseiam nos momentos finais dessas frases. O tom médio total usa da equação (1) de forma continuada durante todo o discurso, enquanto o tom médio se estabelece especialmente para cada frase entoacional. Na medida em que o tom médio se diferencia do tom médio total pelo fato de que o tom médio total não sofre interferências de fronteiras frasais, ele serve como uma referência fixa para a avaliação do tom médio, definido para cada frase entoacional. Assim, havendo uma declinação frasal bem definida ou muitas inflexões, a correlação entre tom médio total e tom médio será muito baixa ou inexistente. Uma correlação negativa também mostra diferenças marcadas entre eles.

Assim, entende-se que a ocorrência de correlação forte entre tom médio e tom médio total é a forma menos marcada para a definição de uma frase assertiva. Relação semelhante pode ser definida pela razão entre tom médio e tom médio total. Na medida em que essa razão se afaste da igualdade, que seria uma razão = 1, entende-se que houve uma tendência maior a uma mudança de registro em relação ao que era esperado. A diferença em relação à variação do tom médio demonstra, dentre outras possibilidades, variação de manifestação emocional na fala. Essa variação de registro emocional já foi verificada em outras pesquisas no âmbito do projeto ExProsodia (VASSOLER; MARTINS, 2013; FERREIRA NETTO *et al.*, 2014).

Nesse caso, ainda que esse critério também se defina em relação às variações de registro frasal no período, entende-se que, havendo variação, essa seria a forma marcada. A dispersão de entoação acima do limite superior do tom médio também tem de ser caracterizada como uma tendência à mudança de registro frasal. Essa dispersão também tem de ser definida a partir da comparação com as variações de registro frasal no período. Em trabalho relativo a gênero (FERREIRA NETTO *et al.*, 2010; FERREIRA NETTO *et al.*, 2016), foi possível verificar que a variação média das frequências acima do tom médio tem caráter significativo e também pode apresentar-se como critério diferenciador de frases entoacionais num mesmo período. Assim, apesar de esse também ser um aspecto pouco

explorado na prosódia do português falado no Brasil, entende-se que uma variação média maior do que 3 semitons de foco/ênfase é mais marcada do que uma variação média menor do que essa.

Uma vez que se tenham estabelecido os momentos limiares de segmentação da série temporal, pode-se questionar sua relação com outras possibilidades de segmentação que envolvam critérios sintáticos e semânticos, por exemplo. No âmbito do programa ExProsodia de pesquisa, a relação entre esses dois aspectos da linguagem, prosódia e léxico, foi discutida anteriormente (FERREIRA NETTO, 2017a, 2018) e se reporta à concatenação de dois sistemas sonoros diferentes. O sistema prosódico tem fonte contínua de alimentação que se relaciona diretamente com necessidades expressivas do falante, envolvendo, por exemplo, seu estado emocional; o sistema lexical tem fonte descontínua de alimentação que se relaciona diretamente com as seleções lexicais do falante, envolvendo instâncias especialmente linguísticas da cognição. De maneira geral, essa proposta segue o princípio estabelecido por Shannon e Weaver (1949) para a comunicação em geral e, bem posteriormente, por Taylor (2000), quando propôs que a fala pudesse ser tomada como uma sequência de eventos entoacionais, que ele chamou de *intonational stream* conectada com uma sequência de eventos segmentais, que ele chamou de *segmental stream*. A concatenação desses sistemas envolveria várias instâncias da análise automática. Neste artigo, é nossa preocupação a interpretação das segmentações interfrasais liminares.

Conforme o levantamento que se apresentou no início deste artigo, a segmentação de frases entoacionais se estabelece por meio de diferentes critérios, quer sejam interrupções no fluxo de informações sonoras, variações de frequência pontuais ou contínuas, ascendentes ou descendentes, duração segmental, com quaisquer combinações entre eles. Mesmo em se tratando de descrições de sentenças assertivas, não há um critério de segmentação que, por si só, estabeleça uma marca de finalização sentencial, mas, pode-se propor que o conjunto das interrupções de fala, variações de frequência, alongamento de segmento final e declinação forme um sistema redundante no sistema prosódico na ocorrência de pontos de finalização frasal.

Segmentação das frases entoacionais: valores dos parâmetros

Em Martins e Ferreira Netto (2018), demonstramos que a segmentação frasal pode ser feita a partir de três parâmetros, a serem usados alternativamente, mas não exclusivos entre si:

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

(i) ocorrência de intervalo ascendente com valor maior do que $F/Esup_{(t)}$ entre os momentos $Z_{(t-1)}$ e $Z_{(t)}$, em que o momento $Z_{(t-1)}$ seja menor ou igual a $F/Einf_{(t)}$;

(ii) interrupção do Ritmo Tonal com duração maior ou igual ao menor valor definido para pausa seguida de um intervalo tonal maior ou igual a $F/Esup_{(t)}$ ou menor ou igual a $F/Einf_{(t)}$;

(iii) interrupção do Ritmo Tonal com duração maior ou igual ao dobro da duração definida para uma pausa.

Um intervalo ascendente de f_0 (*shift*) com valor maior do que $F/Esup_{(t)}$ entre dois momentos consecutivos na série temporal de f_0 associa-se ao fato de que os limites do Tom Médio são, segundo o modelo ExProsodia, os limites para a percepção da variação tonal quando há cadeia segmental (léxico) presente. Aparentemente, a distância de 3 semitons entre uma UBI final antecedendo uma UBI inicial em fronteiras frasais não é suficiente para caracterização de uma segmentação, pois esses intervalos ocorrem com muita frequência na fala espontânea. Dessa maneira, houve a opção por conjugar esse parâmetro com o fato de que o *shift* deveria partir de um tom perceptivelmente grave passível de definir uma finalização autêntica.

As fronteiras frasais, entretanto, podem ocorrer das mais diferentes formas, de tal maneira que, não necessariamente, o recurso a uma finalização autêntica seguida de um $shift \geq F/Esup$ pode representá-las. Por conta disso, optou-se por se estabelecer mais um parâmetro, em que a interrupção do Ritmo Tonal fosse considerada. O fato de haver uma interrupção dessa natureza tem de ser considerada porque o Ritmo Tonal faz parte da estruturação da entoação frasal, estabelecendo o ajuste entre os sistemas prosódico e lexical. Na medida em que esse ajuste é dinâmico, pressupondo um monitoramento constante por parte do locutor, que deve retomar o tom a cada quebra de sonoridade imposta pela cadeia fonológica agregada ao léxico, sua interrupção desencadeia uma possível perda de continuidade nesse monitoramento. Trata-se, portanto, do que se definiu atrás como *pitch reset*, ainda que a declinação frasal não esteja diretamente envolvida, mas especialmente o peso do multiplicador $(t-1)$ e o denominador (t) do Tom Médio, tal como vai na equação (1). Ainda que essa interrupção seja capaz de estabelecer uma fronteira interfrasal, deve-se salientar que, uma vez que o Ritmo Tonal pressupõe o ajuste entre os sistemas, o próprio Ritmo Tonal tem como parte de sua coesão a própria cadeia lexical que impõe, por sua vez, um sentido para a frase em questão. Por conta desse mecanismo coesivo complementar à entoação, entende-se necessário outro parâmetro que se permita implementar como fronteira interfrasal. Nesse caso, o *shift* pode ser tomado como uma marca de início de frase, desde que se o considere tanto em sua forma perceptivelmente

ascendente ou descendente, isto é, com valor maior do que $F/Esup$ ou menor do que $F/Einf$. Nesse caso, entende-se que após uma pausa deverá ocorrer um *shift* perceptível para marcar uma fronteira interfrasal. Esse fenômeno também foi descrito para o russo relativo à segmentação automática (KOCHAROV; KACHKOVSKAIA; SKRELIN, 2019).

Resultados preliminares

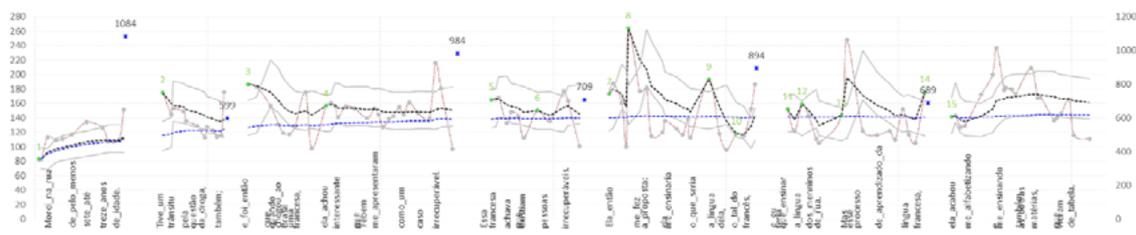
Os dados a serem analisados fazem parte do acervo de dados já coletados no âmbito do projeto ExProsodia. Os arquivos foram transcritos e processados pelo *software* Speech Filing System (HUCKVALE *et al.*, 1987; HUCKVALE, 2008) e pelo aplicativo ExProsodia (FERREIRA NETTO, 2017b, 2010). Os dados acústicos coletados foram tabulados, considerando as variáveis descritas acima. Como ilustração para a aplicação automática desses critérios, analisamos uma narrativa com cerca de 31 s. de duração, a qual pode ser vista no Quadro 1, abaixo.

Quadro 1. Narrativa do sujeito 1, letrado, com fôrma prosódica livre

₁[Morei na rua de pelo menos sete até treze anos de idade.],₂[Tive um trânsito pela questão da droga, também;],₃[e foi então que, quando chegou Brasil uma francesa,],₄[ela achou interessante que na Febem me apresentaram como um caso irre recuperável.],₅[Essa francesa achava que não existiam],₆[pessoas irre recuperáveis.],₇[Ela então],₈[me fez a proposta: ela me ensinaria o que seria],₉[a língua dela,],₁₀[o tal do francês,],₁₁[e eu teria que ensinar],₁₂[a língua dos meninos de rua.],₁₃[Mas esse processo de aprendizado da língua francesa,],₁₄[*pausa ruidosa*],₁₅[ela acabou me alfabetizando e me ensinando também as outras matérias, que vieram de tabela.],₁₅

Fonte: Elaboração própria

Figura 1. Curvas entoacionais referente ao exemplo 1



Fonte: Elaboração própria

O gráfico que vai na Figura 1 foi construído pelo aplicativo ExProsodia. Ele representa a decomposição da curva entoacional de f_0 . A curva entoacional de f_0 vai

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

marcada pela linha cinza pontilhada na qual cada marcador (pequenas bolinhas da mesma cor cinza) representa uma UBI tomada no momento $Z_{(t)}$. A linha preta pontilhada representa o tom médio com os respectivos momentos em que houve *pitch reset*. As linhas pretas contínuas representam, acima e abaixo do tom médio com *pitch reset*, seus limites superiores e inferiores, isto é, variações de 1,19 e 0,84, respectivamente. A linha pontilhada azul representa o tom médio total, definido pela média aritmética cumulativa dos valores de f_0 . As interrupções das linhas representam as pausas, que, por sua vez, têm seus valores marcados em milissegundos acima dos marcadores azuis com forma de cruz. Na Tabela 2, vão os valores obtidos pelo aplicativo ExProsodia. Esses valores foram utilizados para as marcações manuais feitas no exemplo textual do exemplo. Essa tabela é fornecida automaticamente pelo aplicativo.

Tabela 2. Variações de finalização entoacional referente ao exemplo 1

Frase Entoacional	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Direção	0	0	1	1	1	1	1	1	1	0	1	1	1	0	1
Movimento	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Alvo	0	0	1	1	1	1	1	1	1	0	0	1	1	0	1
Alongamento	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
Correlação	0	1	1	0	0	0	1	1	1	1	1	1	1	0	1
Diferença	0	1	1	0	0	0	1	1	0	0	0	0	0	1	0
F/E	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0
Pausa>0,05	1	1	0	1	0	1	0	0	0	1	0	1	0	1	1
Shift	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0

Fonte: Elaboração própria

A partir dos princípios de segmentação automática, estabeleceram-se 15 frases entoacionais, marcadas com um índice no exemplo 1. As frases entoacionais segmentadas e numeradas de 1 a 4 e de 13 a 15, excetuando-se a 14 que ocorreu devido à presença de ruído, apresentaram correspondência em relação à completude sintático-lexical das sentenças. As frases entoacionais 5 a 6, 7 a 10, 11 a 12, entretanto, não apresentaram a mesma correspondência: houve 8 frases entoacionais e apenas três sentenças sintático-lexicais.

As fronteiras entre as frases entoacionais 3 e 4, 5 e 6, 8 e 9, 10 e 11, 12 e 13 estabeleceram-se com base no procedimento I da segmentação automática liminar. Esse procedimento leva em consideração especialmente o valor de *shift* maior ou igual ao de $F/Esup$, a partir do momento anterior de f_0 que tenha ocorrido com alvo autêntico. Essa segmentação liminar ocorreu entre constituintes sintáticos nas sentenças 5-6 e 7-10.

5[Essa francesa achava que não existiam]5 6[peças irreparáveis.]6

7[Ela então]7 8[me fez a proposta: ela me ensinaria o que seria]8 9[a língua dela,]9
10[o tal do francês,]10

Como se pode notar, a segmentação automática liminar, pese ter ocorrido entre constituintes sentenciais, não ocorreu internamente a esses constituintes. De certa maneira, as fronteiras estabelecidas podem ser justificadas, caso fosse necessário, recorrendo-se ao sentido que os trechos segmentados portariam. A frase 7 apresenta uma junção lexicalizada *então*. Esse tipo de junção caracteriza uma fronteira entre unidades sentenciais de narrativas (LABOV, 1997).

A fronteira que se estabeleceu na sentença 5-6, separando os constituintes verbo e objeto, permite-nos supor que a finalização descendente com alvo autêntico da frase 5 caracteriza uma ênfase no intervalo entoacional que virá a seguir, salientando a inexistência de “peças irreparáveis”.

As fronteiras entre 8 e 9, 9 e 10, na sentença 7-10, parecem ter ocorrido devido ao envolvimento de fatos pragmáticos. No caso particular da fronteira entre as frases entoacionais 8 e 9, a finalização ascendente por salto com alvo plagal, que caracteriza um *shift* e, portanto, um *pitch reset*, seria mais propriamente caracterizada por uma marcação de *F/Esup*. Já a segmentação liminar entre as frases entoacionais 9 e 10 pode se justificar pela ocorrência de um acréscimo sintático, isto é, um constituinte recorrente após uma finalização.

Explicação semelhante poderia ser dada à segmentação intrasentencial pelas frases entoacionais 11 e 12. Trata-se de fato pragmático semelhante ao que ocorreu entre as frases entoacionais 9 e 10. A ênfase prevista no objeto vai antecipada pelo estabelecimento de um *F/Einf*e, assim, pode estabelecer uma finalização ascendente por salto que corresponde a uma marca de fronteira interfrasal de entoação.

Em alguns pontos, por sua vez, houve fenômenos entoacionais que, aparentemente, não foram capturados pelos procedimentos ora propostos de segmentação automática liminar. Na sentença 7-10, durante a frase entoacional 8, há uma curva descendente oblíqua autêntica que se segue de uma curva ascendente oblíqua plagal que poderia caracterizar uma fronteira interfrasal de entoação associada, em cada uma das frases entoacionais segmentáveis, a um sintagma intrasentencial. Para que essa captura fosse possível, teria sido necessário que a calibragem inicial para as variações laterais máxima do tom médio fosse menor do que os três semitons assumidos por ora. Embora isso seja possível, e vai previsto no *software* ExProsodia, uma mudança dessa natureza envolveria todos os procedimentos de segmentação automática realizados nessa locução.

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

Os processos de segmentação liminar das frases entoacionais, principalmente, atuam no sentido de efetivar o *pitch reset* para uma nova frase entoacional. Essa renovação promove o que podemos interpretar como uma independência entre as séries temporais do tom médio extraído da elocução tomada globalmente e a do tom médio de cada frase todas localmente. Essa independência entre o tom médio total e o tom médio pode assumir-se como um conjunto binário de vozes autônomas cuja relação produz sentidos suplementares aos sentidos sentenciais.

Na medida em que as curvas decorrentes desses tons médios se mantenham semelhantes, entende-se que não pode haver interpretação diferenciada entre elas. É o caso que se observa na frase 1. Ao se distanciarem, entretanto, é possível perceber acréscimo de sentido. Nas frases 1, 4, 5 e 6, o parâmetro *correlação/dispersão* aponta para a ausência de variações do tom médio local. Conjugado com o parâmetro *diferença*, que apresenta resultados semelhantes, podemos interpretar que essas frases entoacionais não foram objeto de informação marcada. Nas frases entoacionais 9, 10, 11, 12, 13 e 15, o parâmetro *diferença* não mostra informação marcada. Assim, por se considerar que o parâmetro *diferença* aponta para uma intenção de manifestar um estado emocional, podemos tomar que a entoação, nesse caso, não apresentou essa informação, mantendo-se somente dispersa em torno do tom médio local. Essa interpretação se reforça se observarmos que o parâmetro *F/E* para as frases entoacionais 8, 10 e 11 também não mostrou informação marcada, isto é, suas oscilações não ultrapassaram os limites laterais definidos para o tom médio local de cada frase entoacional. Desse ponto de vista, podemos interpretar que somente a frase entoacional 8 apresentou informação marcada para esses três parâmetros – *correlação/dispersão*, *diferença* e *F/E* – tomados conjuntamente. Trata-se de uma frase entoacional, conforme se vê na Tabela 2, que apresenta o maior número de parâmetros com informação marcada. Ainda em relação a essa frase entoacional 8, podemos entender que os parâmetros que caracterizam sua finalização – *direção*, *movimento* e *alvo* – estabelecem que se trata de uma marcação entoacional própria a uma afirmação peremptória bastante marcada.

Considerações finais

Nossos objetivos primários neste artigo eram demonstrar a capacidade do modelo ExProsodia em segmentar automaticamente unidades entoacionais, a partir de parâmetros acústicos amplamente citados na literatura. Em paralelo a isso, nossa intenção é apresentar uma leitura teórica capaz de unir a ideia de sentença entoacional com a noção tradicional do termo. No que tange ao primeiro objetivo, nota-se que segmentação automática da curva de f_0 no discurso falado, conforme já vimos, é uma das preocupações da análise

linguística instrumental desde muitos anos. Para a língua portuguesa falada no Brasil, as proposições de Barbosa e seus colegas (BARBOSA, 2013; MITTMANN; BARBOSA, 2016; TEIXEIRA; BARBOSA; RASO, 2018, dentre outras), que partem de princípios diferentes da que se discutiu aqui, mostraram-se muito promissoras. A proposta de fazê-la pelos critérios adotados pelo modelo ExProsodia, mediante o uso de *software* desenvolvido especificamente para esse fim, mostrou-se como mais uma possibilidade concreta e bem-sucedida. Embora algumas dificuldades ainda tenham de ser ultrapassadas, tais como, pausas preenchidas, alterações *ad hoc* dos limites laterais do tom médio local ou variações na taxa de elocução, ainda assim os avanços que o modelo mostrou em sua aplicação têm de ser considerados positivamente.

Um ponto a ser debatido e reanalisado é o fato de o modelo apresentar segmentações de frases entoacionais em pontos que não correspondem exatamente aos pontos esperados por uma análise puramente sintático-lexical, o que a nosso ver é uma questão que exige uma interpretação mais aprofundada, seja dos parâmetros utilizados na análise, seja das noções de sentença sintático-lexical e de frase entoacional. Conforme já se discutiu alhures, a objetividade da análise instrumental automática não permite qualquer tipo de construção subjetiva sobre estas unidades estruturais. Assim, os resultados obtidos por esse tipo de análise, para que sejam interpretados linguisticamente, têm de receber um tratamento adequado às necessidades humanas compartilhadas, por sua vez, por um grupo de falantes da mesma língua.

A conversão de dados obtidos por aparelhos em informação adaptada às necessidades linguísticas, entretanto, não garante que os dados convertidos sejam precisamente os que a intuição do falante pressupõe para eles. Na medida em que não há como fazer comparações diretas e completamente inequívocas entre intuição humana e análise instrumental automática, deve-se estabelecer que as diferenças decorram de princípios outros que não sejam os da observação introspectiva da linguagem. Essa relação entre essas duas dimensões leva-nos ao segundo objetivo deste artigo.

O modelo ExProsodia pressupõe que haja dois sistemas concorrentes na formação da fala: o sistema sintático-lexical e o sistema prosódico/entoacional. Ainda que ambos atuem sobre a mesma produção sonora da fonação, são sistemas em grande parte independentes que exigem grande esforço do falante para estarem devidamente sincronizados. O sistema sintático-lexical tem sua sonoridade manifesta na cadeia fonológica, cuja principal característica é definida pela variação de timbre com distribuição regulada por regras historicamente definidas. O sistema prosódico/entoacional tem sua sonoridade manifesta especialmente nos momentos da cadeia fonológica em que há segmentos soantes com sonorização espontânea (CHOMSKY; HALLE, 1968; HALLE; STEVENS, 1971).

Ainda que esse amálgama de sistemas crie vínculos fortes entre eles, a geração e o propósito de cada um deles continuam independentes. O sistema prosódico/entoacional, por sua vez, pode ser tomado de um ponto de vista global, em que se sujeita às condições emocionais ou atitudinais do falante e, de um ponto de vista local, em que se sujeita às necessidades estruturais da linguagem, como propõe Pierrehumbert (2000) e Xu (2011). Qualquer que seja o ponto de vista, global ou local, o sistema prosódico/entoacional tem pouca saliência de sentido perceptível em relação ao sistema sintático-lexical (PERES; CONSONI; FERREIRA NETTO, 2011; MADSEN; WHITEFORD; OXENHAM, 2017). Assim, embora todas as variações entoacionais sejam de responsabilidade exclusiva do falante, não há como garantir sua consciência ou sua intencionalidade. Assim, a intuição do falante nesse caso fica prejudicada e não constitui mais um mecanismo seguro de avaliação. Novamente, Luhmann (2016, p. 60) chamou a atenção para isso:

O ser humano pode parecer para si mesmo ou para observadores como unidade, mas ele não é um sistema. Tampouco pode se formar um sistema com um grande número de seres humanos. Com base em tais suposições não se perceberia que o ser humano nem mesmo pode observar em si mesmo aquilo que nele se passa na qualidade de processos fisiológicos, químicos e biológicos. Sua vida é inacessível ao seu sistema psíquico; ele tem de coçar, sentir dor ou, mediante outro meio, chamar a atenção para si, a fim de estimular a operação de um outro nível de formação sistêmica, a consciência do sistema psíquico.

Apesar da possibilidade de comunicação entre sistemas ser apenas parcial, a sujeição do sistema prosódico/entoacional ao sistema sintático-lexical decorre de este último ser o único capaz de gerar interpretantes peircianos (FERREIRA NETTO, 2018). Desse ponto de vista, a entoação atua tanto a partir das necessidades emocionais geradas de processos fisiológicos, químico e biológico, como sugeriu Luhmann, como a partir de simulações desses processos estimulados pela própria semiótica sintático-lexical do discurso. Por conta dessa propriedade, o sistema prosódico/entoacional, apesar de independente, tem de adequar-se constantemente às formas da linguagem e do meio. Essa adequação aponta para a necessidade de se buscar e compreender com maior minúcia as regularidades que ela exige, tanto do ponto de vista dos gêneros discursivos, como do ponto de vista das condições emocionais do falante durante a produção do discurso. Por fim, entendemos que uma discussão mais ampla sobre a natureza destes sistemas e a sua relação com a uso da linguagem é um tema a ser explorado com mais profundidade em pesquisas subsequentes.

Referências

- BANSE, R.; SHERER, K. R. Acoustic Profiles in Vocal Emotion Expression. **Journal of Personality and Social Psychology**, v. 70, n. 3, p. 614-636, 1996.
- BARBOSA, P. A. Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. TRASP'2013, Aix-en-Provence. *In*: BIGI, B.; HIRST, D. (ed.). **Proceedings...**, Aix-en-Provence, 2013. p. 86-90.
- BAZ, D. G. M. **As relações entre entoação frasal e melodia de músicas populares paraguaias**. 2011. Tese (Doutorado em Letras Clássicas e Vernáculas) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.
- BOLINGER, D. **Intonations and its Uses. Melody in grammar and discourse**. Stanford: Stanford University Press, 1989.
- BOOMER, D. S.; DITTMANN, A. T. Hesitation pauses and juncture pauses in speech. **Language and Speech**, v. 5, n. 4, p. 215-220, 1962.
- CAGLIARI, L. C. **Elementos de Fonética do Português Brasileiro**. 1981. Tese (Livre-Docência em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1981.
- CAMPBELL, N. Automatic detection of prosodic boundaries in speech. **Speech Communication**, v. 13, n. 3-4, p. 343-354, dez. 1993.
- CHOMSKY, N.; HALLE, M. **The Sound Patterns of English**. New York: Harper & Row, 1968.
- COHEN, A.; HART, J. T. Anatomy of intonation. **Lingua**, v. 19, n. 2, p. 177-192, 1967.
- CONSONI, F. **Aspectos da percepção da proeminência tonal em português brasileiro**. 2011. Tese (Doutorado em Letras Clássicas e Vernáculas) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.
- COOPER, W. E.; SORENSEN, J. M. **Fundamental Frequency in Sentence Production**. New York: Springer-Verlag, 1981.

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

COOPER, W. E.; SORENSEN, J. M. Fundamental frequency contours at syntactic boundaries. **Journal of the Acoustical Society of America**, v. 62, n. 3, p. 683-692, 1977.

DANEŠ, F. Sentence intonation from a functional point of view. **Word-Journal of the International Linguistic Association**, v. 16, n. 1, p. 34-54, 1960.

DARWIN, C. **A expressão da emoção no homem e nos animais**. São Paulo: Companhia das Letras, 2000.

DELATTRE, P. Comparison of syllable length conditioning among languages. **Iral-International Review of Applied Linguistics in Language Teaching**, v. 4, n. 3, p. 183-198, 1966.

DI CRISTO, A. L'intonation est congruente à la syntaxe: une confirmation. *In*: ROSSI, A. D. C. M.; HIRST, D.; MARTIN, P.; NISHINUMA, Y. (ed.). **L'Intonation, de la acoustique a la sémantique**. Paris: Klincksieck, 1981. p. 272-289.

DITTMANN, A. T.; LLEWELLYN, L. G. Phonemic clause as a unit of speech decoding. **Journal of Personality and Social Psychology**, v. 6, n. 3, p. 341-351, 1967.

DUEZ, D. Perception of silent pauses in continuous speech. **Language and Speech**, v. 28, p. 377-389, out./dez. 1985.

FAIRBANKS, G. Recent experimental investigations of vocal pitch in speech. **Journal of the Acoustical Society of America**, v. 11, n. 4, p. 457-466, abr. 1940.

FERREIRA NETTO, W. Narrativa oral e prosódia. **Policromias – Revista de Estudos do discurso, imagem e som**, v. 3, n. 1, p. 48-68, 2018.

FERREIRA NETTO, W. Fonologia, Discurso e Interdisciplinaridade. Relações possíveis entre a fonologia e a música. **SEMINÁRIO INTERNACIONAL DE LINGUÍSTICA**. São Paulo 2017a.

FERREIRA NETTO, W. O programa ExProsodia de análise automática da entoação: aplicações e resultados. SIMPÓSIO INTERNACIONAL SOBRE LINGUAGEM E COGNIÇÃO: FMRI APLICADA A PESQUISAS EM LINGUÍSTICA COGNITIVA. LIMA-HERNANDES, M. C. São Paulo 2017b.

FERREIRA NETTO, W. Análise automática de manifestações emocionais em PB: aplicações do programa ExProsodia. *In*: FERREIRA NETTO, W. (org.). **ExProsodia**. Resultados Preliminares. São Paulo: Paulistana, 2016. p. 11-28.

FERREIRA NETTO, W. Finalização de frases na música e na fala. SIMPÓSIO EM NEUROCIÊNCIA E COGNIÇÃO MUSICAL. São Bernardo 2015.

FERREIRA NETTO, W. ExProsodia. **Revista de Propriedade Industrial**. PAULO, U. D. S. Brasil. RS 08992-2 2010.

FERREIRA NETTO, W.; PERES, D. O.; CONSONI, F. Relações entre variação de gênero e variação tonal na fala de língua portuguesa do Brasil. **Revista Symposium (Lavras)**, v. 7, n. 1, p. 70-73, 2010.

FERREIRA NETTO, W.; PERES, D. O.; MARTINS, M. V. M. Tentativa de disposição de registros entoacionais num eixo horizontal organizado pela tensão entoacional. **Gradus**, v. 2, n. 1, p. 14-29, 2017.

FERREIRA NETTO, W.; CONSONI, F. Estratégias prosódicas na leitura em voz alta e da fala espontânea. **Alfa**, Revista de Linguística, v. 52, n. 2, p. 521-534, 2008.

FERREIRA NETTO, W.; CONSONI, F.; PERES, D. O. Finalizações de frases em leituras e frases espontâneas no PB. SEMINÁRIO DO GEL, 57. Ribeirão Preto: Grupo de Estudos Linguísticos do Estado de São Paulo 2009.

FERREIRA NETTO, W.; MARTINS, M. V. M.; VIEIRA, M. D. F. Efeitos da entoação e da duração na análise automática das manifestações emocionais. **Estudos Linguísticos**, v. 43, p. 22-32, 2014.

FERREIRA NETTO, W.; PERES, D. O.; MOREIRA, M. M. V. Comparação entre o Tom Médio e a Moda tomados numa banda de 7 semitons como referência para variações entoacionais. ENCONTRO NACIONAL DA ANPOLL, 31. Campinas, 2016.

FLETCHER, J. Some micro and macro effects of tempo change on timing in French. **Linguistics**, v. 25, n. 5, p. 951-967, 1987.

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

FRIEDMAN, L. A.; OCONNELL, D. C. PAUSE REPORTS FOR SPONTANEOUS DIALOGIC SPEECH. **Bulletin of the Psychonomic Society**, v. 29, n. 3, p. 223-225, maio 1991.

GARVIN, P. L.; MATHIOT, M. Fused units in prosodic analysis. **Word-Journal of the International Linguistic Association**, v. 14, n. 2-3, p. 178-186, 1958.

HALLE, M.; STEVENS, K. N. A note on laryngeal features. *In*: HALLE, M.; STEVENS, K. N. (ed.). **Quarterly Progress Report (Research Laboratory of Electronics, MIT)**, v. 101, p. 198-213, 1971.

HARGREAVES, W. A.; STARKWEATHER, J. A. Collection of temporal data with the duration tabulator. **Journal of the Experimental Analysis of Behavior**, v. 2, n. 3, p. 179-183, 1959.

HOCKETT, C. F. A system of a descriptive phonology. **Language**, v. 18, n. 1, p. 3-21, 1953.

HUCKVALE, M. **Speech Filing System v.4.7/Windows SFSWin** 2008.

HUCKVALE, M. A.; BROOKES, D. M.; DWORKIN, L. T.; JOHNSON, M. E.; PEARCE, D. J.; WHITAKER, L. C. **The SPAR Speech Filing System**. EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY. Edinburgh, 1987.

JONES, D. **An Outline of English Phonetics**. New York: G. E. Stechert & Co., 1922.

KOCHAROV, D.; KACHKOVSKAIA, T.; SKRELIN, P. Prosodic boundary detection using syntactic and acoustic information. **Computer Speech & Language**, v. 53, p. 231-241, 2019.

KOHLER, K. J.; NIEBUHR, O. The phonetics of emphasis. *In*: TROUVAIN, J.; BARRY, W. J. **International Congress of Phonetic Sciences (ICPhS)**, 2007, Saarbrücken. Jürgen Trouvain; William J. Barry. p. 2145-2148.

KUHL, P. K. Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. **Perception & Psychophysics**, v. 50, n. 2, p. 93-107, ago. 1991.

LABOV, W. Some further steps in narrative analysis. **The Journal of Narrative and Life History**, v. 7, n. 1-4, p. 395-415, 1997.

LADD, R. Declination: a review and some hypotheses. **Phonology Yearbook**, v. 1, p. 53-74, 1984.

LIEBERMAN, P. **Intonation, Perception, and Language**. Cambridge: The M.I.T. Press, 1967.

LINDBLOM, B. E. F. Temporal organization of syllable production. **Quarterly Progress and Status Report**, v. 9, n. 2-3, p. 1-5, 1968.

LUHMANN, N. **Sistemas Sociais: esboço de uma teoria geral**. Petrópolis: Vozes, 2016.

LÖVGREN, T.; DOORN, J. V. Influence of manipulation of short silent pause duration on speech fluency. DiSS'05, **Disfluency in Spontaneous Speech Workshop**, 2005, Aix-en-Provence. ISCA Archive. p. 123-126.

MADSEN, S. M. K.; WHITEFORD, K. L.; OXENHAM, A. J. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. **Scientific Reports**, v. 7, Oct 3 2017.

MAEDA, S. **A characterization of American English intonation**. 1976. Electrical Engineering, Massachusetts Institute of Technology, Massachusetts, 1976.

MARTINS, M. V. M.; FERREIRA-NETTO, W. Prosodic Boundaries and Automatic Intonation Analysis. 6th EICEFALA PROCEEDINGS, 2018. Belo Horizonte. p. 16-17.

MARTINS, M. V. M.; FERREIRA NETTO, W. Os limiars de diferenciação tonal do português brasileiro. **Revista do GEL**, v. 14, n. 2, p. 157-182, 2017.

MITTMANN, M. M.; BARBOSA, P. A. An automatic speech segmentation tool based on multiple acoustic parameters. **CHIMERA**. Romance Corpora and Linguistic Studies, v. 3, n. 2, p. 133-147, 2016.

MORAES, J. A. Intonation in Brazilian Portuguese. *In*: HIRST, D.; CRISTO, A. D. (ed.). **Intonation System. A survey of twenty languages**. Cambridge: Cambridge, 1998. Cap. 10, p. 1749-194.

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

MORAES, J. A. D. **Intonational Phonology of Brazilian Portuguese**. WORKSHOP ON INTONATIONAL PHONOLOGY: UNDERSTUDIED OR FIELDWORK LANGUAGES, ICPHS 2007 SATELLITE MEETING. Saarbrücken 2007.

MORAES, J. A.; RILLIARD, A. Describing the intonation of speech acts in Brazilian Portuguese: Methodological aspects. *In*: FELDHAUSEN, J. F.; VANRELL, M. del M. (ed.). **Methods in prosody: A Romance language perspective**. Berlin: Language Science Press, 2018. p. 229-262. DOI: <https://doi.org/10.5281/zenodo.1441347>

OLIVEIRA JR, M. Pitch reset as a cue for narrative segmentation. *In*: AMINA METTOUCHI, G. F. Nantes: Interfaces Prosodiques/Prosodic Interfaces, 2003. p. 73-79.

OLIVEIRA, M. The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. *In*: RANCHHOD, E.; MAMEDE, N. J. (ed.). **Advances in natural language processing: third international conference; proceedings**. Berlin: Springer, 2002. p. 43-51.

OLLER, D. K. Effect of position in utterance on speech segment duration in English. **Journal of the Acoustical Society of America**, v. 54, n. 5, p. 1235-1247, 1973.

OHALA, J. J. An Ethological Perspective on Common Cross-Language Utilization of Fo of Voice. **Phonetica**, v. 41, n. 0, p. 1-16, 1984.

PERES, D. O.; CONSONI, F.; FERREIRA NETTO, W. A influência da cadeia segmental na percepção de variações tonais. **LLJournal**, v. 6, n. 1, p. 1-13, 2011.

PIERREHUMBERT, J. B. Automatic recognition of intonation patterns. *In*: MARCUS, M., ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 1983. Cambridge. Association for Computational Linguistics. p. 85-90.

PIERREHUMBERT, J. B. **The Phonology and Phonetics of English Intonation**. Cambridge: Linguistics and Philosophy, Massachusetts Institute of Technology, 1980.

PIERREHUMBERT, J. B. Tonal Elements and their Alignment. *In*: HORNE, M. (org.). **Prosody, Theory and Experiment: studies presented to Gösta Bruce**. Dodrecht: Kluwer Academic Publisher, 2000. p. 11-36.

PIKE, K. L. **The Intonation of American English**. Ann Arbor: The University of Michigan Press, 1945.

PRICE, P. J.; OSTENDORF, M.; WIGHTMAN, C. W. Prosody and parsing. *In: LINGUISTICS, A. F. C. WORKSHOP ON SPEECH AND NATURAL LANGUAGE – HLT, 1989, Cape Cod.* p. 5-11.

PRONOVOST, W. **An experimental study of the habitual and natural pitch levels of superior speakers**. 1939. Ph.D. State University of Iowa, Iowa, 1939.

ROSA, R. C. D. M. **Percepção e ExProsodia: correlação entre análise automática e a finalização de frases assertivas isoladas do Português Brasileiro**. 2015. Dissertação (Mestrado em Letras Clássicas e Vernáculas) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

ROSA, R.; CONSONI, F.; FERREIRA NETTO, W. A questão da correlação entre a análise automática das finalizações prosódicas e a separação intuitiva de frases em textos longos. *In: FERREIRA NETTO, W. (org.). ExProsodia: resultados preliminares*. São Paulo: Paulistana, 2016. p. 61-63.

SCHERER, K. R. Expression of Emotion in Voice and Music. **Journal of Voice**, v. 9, n. 3, p. 235-248, 1995.

SCHOENBERG, A. **Exercícios preliminares em contraponto**. São Paulo: Via Lettera, 2001a.

SCHOENBERG, A. **Harmonia**. São Paulo: Editora UNESP, 2001b.

SHANNON, C. E.; WEAVER, W. **The Mathematical Theory of Communication**. Urbana: The University of Illinois, 1949.

SILVA, E. W. R. D. **A relação entre produção e percepção de pistas prosódicas na segmentação de narrativas espontâneas**. 2017. Dissertação (Mestrado em Letras) – Universidade Federal de Alagoas, Maceió, 2017.

SNIDECOR, J. C. The Pitch And Duration Characteristics Of Superior Female Speakers During Oral Reading. **Journal of Speech and Hearing Disorders**, v. 16, n. 1, p. 44-52, mar. 1951.

- | Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto

SNIDECOR, J. C. An objective study of phrasing in impromptu speaking and oral reading. **Speech Monographs**, v. 11, p. 97-104, 1944.

SNIDECOR, J. C. A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. **Speech Monographs**, v. 10, p. 50-56, 1943.

SPENCER, H. The origin of music. **MInd**, p. 449-468, 1890.

STETSON, R. H. **Motor Phonetics. A study of speech movements in action**. Amsterdam: Springer, 1927.

T'HART, J.; COLLIER, R.; COHEN, A. **A perceptual study of intonation. An experimental-phonetic approach to speech melody**. Cambridge: Cambridge University Press, 1990.

TAYLOR, P. Analysis and synthesis of intonation using the Tilt model. **Journal of the Acoustical Society of America**, v. 107, n. 3, p. 1697-1714, mar. 2000.

TAYLOR, P. The rise fall connection model of intonation. **Speech Communication**, v. 15, n. 1-2, p. 169-186, out. 1994.

TEIXEIRA, B.; BARBOSA, P.; RASO, T. Automatic Detection of Prosodic Boundaries in Brazilian Portuguese Spontaneous Speech. *In*: VILLAVICENCIO, A.; MOREIRA, V.; ABAD, A.; CASELI, H.; GAMALLO, P.; RAMISCH, C.; OLIVEIRA, H. C.; PAETZOLD, G. H. (ed.). **Computational Processing of the Portuguese Language**. 13th International Conference, PROPOR 2018 Canela, Brazil, September 24–26, 2018 Proceedings. Cham: Springer, 2018. p. 429-437. DOI: <https://doi.org/10.1007/978-3-319-99722-3>

TRAGER, G. L.; SMITH, H. L. **An Outline of English Structure**. Washington: American Council of Learned Societies, 1957.

TRUBETZKOY, N. S. **Principios de fonologia**. Madrid: Editorial Cincel, 1973.

VAISSIÈRE, J. Language Independent Prosodic Features. *In*: CUTLER, A.; LADD, R. (ed.). **Prosody: Models and Measurements**. Heidelberg: Springer Verlag, 1983. p. 53-65.

VAISSIÈRE, J. **On French prosody**. Massachusetts Institute of Technology. Cambridge, 1974. p. 212-223.

VASSOLER, A. M. O.; MARTINS, M. V. M. A entoação em falas teatrais: uma análise da raiva e da fala neutra. **Revista Estudos Linguísticos**, v. 42, n. 1, p. 9-18, 2013.

VERZEANO, M.; FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. **Science**, v. 110, n. 2845, p. 45-46, 1949.

XU, Y. Speech prosody: A methodological review. **Journal of Speech Sciences**, v. 1, n. 1, p. 85-115, 2011.

COMO CITAR ESTE ARTIGO: NETTO, Waldemar FERREIRA; MARTINS, Marcus Vinícius Moreira. Estratégias de finalizações frasais e a análise automática da entoação: um estudo piloto. **Revista do GEL**, v. 17, n. 3, p. 90-118, 2020. Disponível em: <https://revistadogel.gel.org.br/>

DOI: <http://dx.doi.org/10.21165/gel.v17i3.2849>

Submetido em: 17/06/2020 | Aceito em: 16/09/2020.
