

## Variações entoacionais em narrativas populares: análises preliminares da codificação do contexto

Intonational variations in popular narratives: preliminary analysis of context coding

**Marcus Vinícius Moreira Martins**

Universidade do Estado de Minas Gerais

**Waldemar Ferreira Netto**

Universidade de São Paulo

### Marcus Vinícius Moreira Martins

Professor da Universidade do Estado de Minas Gerais (UEMG), Doutor em Filologia e Língua Portuguesa (2017), Mestre em Filologia e Língua Portuguesa (2013). É membro colaborador do Grupo ExProsodia (USP) e do Laboratório de Fala e Linguagem - Lafalin (USP). <https://orcid.org/0000-0002-6835-5639>

### Waldemar Ferreira Netto

Doutor em Lingüística, na área de educação escolar indígena guarani, pela Universidade de São Paulo (1994), e livre-docência sobre a prosódia da língua portuguesa na Universidade de São Paulo (2006). Atualmente é Professor Titular da Universidade de São Paulo. <http://orcid.org/0000-0002-4136-341X>

Recebido em:  
21/10/2021

Aceito em:  
31/01/2022

JAN / ABR 2022  
ISSN 2317-9945 (on-line)  
ISSN 0103-6858  
p. 118 - 134

### RESUMO

O *software* ExProsodia foi reformulado de maneira a contemplar dois pontos de referência: o tom médio global (o tom médio propriamente dito) e o tom médio local. O tom médio local usa os valores do *reset* frasal para segmentações específicas e reinicia o processo de cálculo da média acumulada no tempo. A hipótese é a de que haja correlação entre cada um desses momentos de reset frasal e a constituição de frases lexicamente consideradas. Esses pontos de referência possibilitam interpretar a entoação para com maior espectro de variações tonais condicionadas intencional e não intencionalmente. Propõe-se que variações do tom médio global não sejam intencionais e que as variações locais sejam intencionais para a expressividade desejada. Para corroborar essa hipótese, foram analisadas 10 narrativas orais populares. As análises usaram a versão atual do software ExProsodia. Os resultados corresponderam parcialmente à hipótese proposta, sobretudo nos momentos de maior tensão narrativa.

### PALAVRAS-CHAVE

Narrativas orais. Entoação. Português Brasileiro.

### ABSTRACT

The ExProsodia software has been reformulated to include two reference points: the overall mean tone (average accumulated in time) and the local mean tone. The local mean tone uses the phrasal reset values for segmentations and resets the mean tone. The hypothesis is that there is correlation between each of these moments of phrasal reset and the constitution of lexically considered sentences. These points of reference make it possible to interpret the intonation for a larger spectrum of tonal variations conditioned intentionally and unintentionally. It is proposed that variations of overall mean tone are not intentional and that local variations are intentional to the desired expression. To corroborate this hypothesis, 10 popular oral

narratives were analyzed. The analyzes used the current version of ExProsodia software. The results corresponded partially to the hypothesis proposed, especially in moments of greater narrative tension.

## KEYWORDS

Oral narratives. Intonation. Brazilian Portuguese.

## 1. Introdução

Bolinger (1958) descrevia a frequência fundamental como uma selvagem meio domada. Em sua leitura, a implementação de  $f_0$ , por um lado, seria organizada por um conjunto limitado de regras, de modo que as funções associadas a essas regras seriam, necessariamente, gramaticalizadas e incorporadas à estrutura discreta da fonologia, por isso, domada. Por outro lado, a frequência fundamental também toma parte na expressividade do falante, nesse caso, a implementação seria ligeiramente arbitrária e mais voluntária, ainda que seja possível encontrar diversos padrões.

Dentro desta perspectiva, entendemos que a *codificação do contexto* em uma elocução está relacionada ao domínio expressivo e pode ser implementada na fala por meio da prosódia, em especial nas narrativas populares, nas quais o falante, muitas vezes, se apropria do contexto para construir seu discurso. Em trabalhos anteriores (FERREIRA NETTO, 2008; MARTINS, 2017), observamos que o falante mantém uma variação melódica global, a qual serve de parâmetro para a implementação das funções estruturais ou variações não intencionais, como aquelas gramaticalizadas ou associadas ao estado emocional do falante<sup>1</sup>. No mais, o aspecto expressivo acima citado estaria associado a segmentos menores, os quais são controlados pelo falante para obter algum efeito de sentido ou expressividade na sua elocução.

Assim, um primeiro objetivo desta pesquisa se define em relação à análise dos recursos entoacionais usados em narrativas populares, uma vez que entendemos ser necessário verificar se a entoação cumpre uma função gramatical nas narrativas, tal como ela se manifesta no Português Brasileiro. Além disso, outro objeto para esta pesquisa é a discriminação dos processos gramaticais particulares do português usado em narrativas populares e associados ao processo de incorporação e correlação do contexto na produção destas narrativas. Para testar essa hipótese faremos a segmentação e a análise automática da entoação de 10 narrativas e cotejar os resultados com a análise proposta em Labov e Waletzky (1966), a fim de avaliar o grau de aderência da tarefa automatizada. Nas seções a seguir apresentaremos uma discussão teórica dos termos incorporação e correlação seguida da análise das narrativas.

## 2. Contexto de situação: correlação e incorporação

Stephen Levinson (2007), ao discriminar o estudo da Pragmática, deu como

---

1 Tal como Damásio (1994), entendemos que os estados emocionais são ativações corpóreas (bioquímicas e físicas), gerados por estímulos externos ou internos, a fim de colocar um determinado organismo em estado de ação, deste modo a emoção é um estado mental/córporeo involuntário, pelo menos em seu onset.

definição que “Pragmática é o estudo entre língua e contexto que são gramaticalizados ou codificados na estrutura de uma língua.” (LEVINSON, 2007, p. 11). Em relação ao contexto, o autor utiliza, dentre outras, a proposição de Jakobson (1969) para “referência”. Jakobson constrói esse conceito, ao recuperar o conceito de *Bedeutung* (FREGE, 1948), tratando-o como *reference*<sup>2</sup>. No entanto, Frege (1948), que havia proposto a noção de *Bedeutung* no final do século XIX, via a referência como uma condição necessária para o estabelecimento da verdade para enunciado.(incoerente)<sup>3</sup>. Por sua vez, a noção de *contexto de situação*, tal como proposta por Malinowski (1923) e retomada por Jakobson (1980), sugere que o contexto envolveria materialidades ambientais, bem como as condições físicas e emocionais dos atores. Desse ponto de vista, podemos notar que a proposição de “referência” assume um caráter diferenciado do que seria o contexto de situação.

Essa ambiguidade terminológica sobre a noção de contexto leva Levinson (2007) a afirmar que “como deixamos de apresentar uma noção clara de contexto, é provável que o que incluímos no contexto seja tudo o que excluímos da semântica, à guisa das relações de significado” (LEVINSON, 2007, p. 29). Por conta da indefinição acerca desse aspecto, crucial para o desenvolvimento de sua proposta de estudo pragmático, sua definição final ficou ligeiramente mais branda. Assim, segundo o autor, “a pragmática é o estudo do papel que o contexto desempenha no significado do falante.” (LEVINSON, 2007, p. 29). Para evitar uma discussão mais ampla a acerca da definição exata de contexto, é possível assumir para este trabalho que a noção de contexto de que o estudo da pragmática levinsoniana se vale é a de *contexto de situação*, proposta por Malinowski (1923).

Dentro desta definição, a codificação do contexto permite inserir não só elementos do ambiente em que a enunciação ocorre, mas também elementos dos próprios atores envolvidos nessa atividade. Como resultado, estados individuais ou elementos presentes no ambiente, poderiam ser codificados na língua. Sobre esta questão, Levinson (2007) propõe que a relação entre língua e contexto se dá por *correlação* ou por *incorporação*. De acordo com o autor, a correlação seria, por exemplo, a fala arrastada de um bêbado, já a incorporação seria a utilização do pronome pessoal “*tu*” em francês como índice de intimidade entre os atores envolvidos numa enunciação.

Ainda que grosseiramente, podemos entender que a noção de *correlação* se estabelece entre um fato linguístico e um evento contextual, a partir de uma relação causal espontânea cuja origem é a condição fisiológica individual do falante. A mesma relação já era proposta por Darwin (2001), no século XIX, que havia conduzido uma descrição bastante detalhada das condições fisiológicas humanas decorrentes das emoções. Darwin (2001) entende que a fala é produzida a partir do próprio corpo, de modo que as alterações fisiológicas causadas pelas emoções atuam diretamente no corpo, produzindo, inclusive, variações correlatas a elas na materialidade sonora da linguagem. Abordagem semelhante é contemplada por Damásio (1994) e por Bechara et al. (2005).

---

2 Em Ferreira Netto (2017), para discussão mais detalhada.

3 cf. Russel, 1978 para uma crítica a essa condição.

Entretanto, porque o ser humano dispõe de grande habilidade cognitiva, simbólica, que o torna um exímio codificador desse contexto de situação, era de se esperar que a interferência da fisiologia das emoções fosse percebida e simulada também de forma simbólica, para que pudesse ser utilizada conforme suas necessidades expressivas. Essa simbolização, que Levinson (2007) chamou de *incorporação*, para ser tomada como fenômeno efetivamente linguístico e, portanto, objeto de estudo da pragmática, teria de diferenciar-se da correlação mediante a aplicação de alguns critérios. A incorporação deve ser: (i) intencional, (ii) convencional, (iii) codificada num conjunto contrastivo de elementos e, finalmente, (iv) deve estar sujeita a processos gramaticais particulares.

Por sua vez, se usarmos os mesmos quatro critérios para a definição da correlação, teremos que a correlação deve apresentar pelo menos um deles marcado negativamente, isto é, a correlação não deve ser intencional, convencional, codificada ou sujeita a processos gramaticais particulares. Para a incorporação, ao contrário, esses critérios devem ser todos positivos, ou seja, nenhum dos critérios poderia ser dispensado. Na correlação, como estamos lidando com critérios alternativos, todos eles podem ser dispensados. Essa interpretação nos permite, pois, estabelecer graus de relação entre a língua e o contexto, sendo que o grau máximo é o da incorporação e o grau mínimo é o de que nenhum dos critérios propostos ocorra positivamente.

Por estarmos lidando com aspectos prosódicos de narrativas populares, temos de definir em que medida a relação entre esses aspectos prosódicos e o contexto é uma incorporação ou uma correlação. Para isso, a aplicação dos critérios acima descritos exige que se tenham bastante claras as noções de “intenção”, “convenção”, “codificação” e “sujeição a processos gramaticais específicos”. Quanto à noção de *intenção*, podemos usar a proposição de Searle (2002; 2010) que entendia haver estados mentais, intencionais ou não, conscientes ou não. De acordo com o autor:

[...] a experiência de agir é uma experiência com um conteúdo Intencional e a intenção em ação é apenas o componente Intencional, esteja ou não contida em alguma experiência consciente de ação. Vez por outra, realizamos ação sem a menor experiência consciente de o fazer; em um tal caso, a ação existe sem nenhuma experiência de ação (SEARLE, 2002, p. 128).

Nessa passagem, Searle (2002) remete a Robert Mead, quando este propôs a noção de “*self* em ação” (MEAD, 1973, p. 136). A despeito, entretanto, da noção de consciência da ação propriamente dita, Searle (2002) enfatiza a noção de Intencionalidade, ou de estado Intencional, caracterizado pela distinção entre o estado em si e o objeto a que se direciona. Ele dá como exemplo a “ansiedade”, que não pode ser caracterizada como um estado Intencional porque ela é o seu próprio estado, não tendo nenhum objeto a que se direcione (SEARLE, 2002, p. 3). Como exemplo de estado Intencional, ele dá uma lista longa, que inclui: crença, temor, desejo, ódio, aversão, ira, perplexidade, afeição, alegria etc. (SEARLE, 2002, p. 5). Por se caracterizar pela presença de um objeto alvo de sua ocorrência, a intenção (com inicial minúscula) pode ser tomada como uma das formas da Intencionalidade (com inicial maiúscula), caracterizando-se pela pretensão consciente,

ou não, de se atingir seu objeto. Desse ponto de vista, o primeiro critério apontado por Levinson (2007) para distinguir incorporação de correlação se caracterizaria pela pretensão de o locutor manifestar seu estado emocional a seu interlocutor.

Quanto à noção de *convenção*, que caracterizaria o segundo critério proposto por Levinson (2007), podemos ir mais ao passado e buscar em Max Weber algum esclarecimento. Weber (2015), a cavaleiro do século XIX para o XX, propôs que as convenções fossem costumes que sofressem a ação coercitiva dos indivíduos sem que essa ação estivesse formalizada como lei. Para essa diferenciação, Weber (2015, p. 21) propôs que a ação coercitiva da sociedade fosse o critério principal. Os costumes apenas sofreriam maiores ou menores incomodidades ou inconveniências, mas não seriam obrigatórios, e as leis seriam regulamentadas por um grupo de pessoas especializadas e formalmente organizadas para esse propósito coercitivo, com direito e dever de aplicação de penalidades. As convenções estariam numa instância intermediária, pois teriam caráter coercitivo e aplicação de penalidades, sem contar, entretanto, com grupos especializados e organizados para essa avaliação. Assim, é de se esperar que, pela definição de Weber (2015), o caráter coercitivo das convenções atue na prosódia de maneira a padronizar comportamentos, cuja falta será objeto de penalização, provavelmente, no caso, na forma de preconceitos e estigmatizações ou da incompreensão de atitudes entre atores.

No que diz respeito à *codificação linguística do contexto*, é possível tomar a noção de elementos componentes autorreferenciais, desenvolvida por Niklas Luhmann, na segunda metade do século XX. Luhmann (2016), ao propor seu modelo de sistemas autorreferenciais, estabeleceu que seus elementos componentes se caracterizariam por serem desontologizados e definidos mutuamente por suas diferenças. Sua proposição diferia daquela dos modelos estruturalistas da primeira metade do século XX (BLOOMFIELD, 1961; SAUSSURE, 1974), dentre outros aspectos, pelo fato de que as inserções de novos elementos teriam de partir do ambiente, mas com experiência diferencial já estabelecida de maneira que pudessem criar as condições necessárias para que o estímulo externo atuasse no sistema acolhedor (LUHMANN, 2016, p. 61-62).

Desta maneira, seria o próprio sistema, por meio da rede de diferenciações formadora do conjunto de seus elementos componentes, que definiria a pertinência ou não da informação nova a ser codificada. Em outras palavras, a nova informação já teria de estar codificada nos termos do próprio sistema acolhedor para que pudesse, então, ser acolhida por ele. No que tange à prosódia, esse critério exige que se tenha um conjunto de elementos diferenciais já estruturados que caracterizem uma rede de relações prévias entre a língua, de um lado, e contexto e atores da enunciação, de outro.

Finalmente, a definição de *processos gramaticais próprios da língua* em questão envolveria um conjunto de elementos que estivessem normativamente regulados por processos cognitivos próprios desse sistema. Desse ponto de vista, a proposta de análise ExProsodia estabeleceu um conjunto de elementos ordenados cuja aplicação parece regulamentar-se para cada língua. Em nossa proposta assumimos que a entoação se estabelece como uma série temporal, cuja decomposição parte de um valor de referência,

chamado de tom médio, o qual é obtido a partir da média móvel dos  $n$  elementos que compõem a curva de  $f_0$  de uma elocução. Em paralelo a esse valor, há os limites laterais, 3 semitons acima e abaixo do tom médio, formando assim uma faixa de 6 semitons. Esta faixa foi determinada experimentalmente, com base nos limiares de diferenciação tonal (CONSONI, 2011; MARTINS, 2017), de tal modo que os valores que ultrapassavam esse intervalo foram discriminados como Foco/Ênfase, acima, e Finalização, abaixo.

A abordagem baseada em séries temporais nos permite obter dois pontos de referência — um tom médio global e o tom médio local — para interpretar a curva entoacional. O primeiro é o tom médio propriamente dito, como descrito acima. O tom médio local toma como ponto de partida para seus cálculos os valores estabelecidos a partir do *reset* frasal. Esse *reset* frasal estabelece segmentações específicas na série temporal das frequências e reinicia o processo de cálculo da média acumulada no tempo. A hipótese desse tom médio local é a de que haja correlação entre cada um desses momentos de *reset* frasal e a constituição de frases lexicamente consideradas. Em outras palavras, temos uma série geral composta de séries menores que estão interrelacionadas.

As referências associadas às segmentações propostas para os *resets* frasais, ou seja, o tom médio local, nos permite obter uma gama maior de dados observáveis, capazes de cobrir igualmente maior espectro de variações tonais condicionadas tanto por necessidades expressivas intencionais, quanto não intencionais. De maneira geral, propõe-se que as variações do tom médio global sejam não intencionais, uma vez que estão associadas às condições emocionais do falante (FERREIRA NETTO et al., 2014), por sua vez as variações do tom médio local sejam controladas pelo falante para obter algum tipo de expressividade desejada. Neste sentido, entende-se que, pelo menos esses parâmetros são utilizados como recursos gramaticalmente previsíveis no âmbito da língua portuguesa para a produção de narrativas populares.

## 2.1. MATERIAIS E MÉTODOS

Para a análise dos tons médios global e local, como processos gramaticais próprios da língua, coletaram-se 10 narrativas no site *Youtube*, em que os locutores eram homens. A seleção se baseou na qualidade da gravação, especialmente quanto à ausência de ruídos de fundo, falas sobrepostas e músicas. Além dos vídeos, foram utilizadas gravações de narrativas do ciclo de Lampião, feitas também com voz masculina, obtidas de pesquisa realizada anteriormente (CONCEIÇÃO et al., 2016). A extração do som dos vídeos foi feita com o aplicativo VLC 3.0.6 Vetinari (VIDEOLAN, 2019). Todos os arquivos sonoros foram processados usando o Adobe Audition 3 (ADOBE, 2012). Alguns arquivos passaram por um filtro *Dinamic EQ* para cobrir os sons indesejados sem perder a duração e a frequência fundamental. Também, em alguns casos, foi necessário eliminar vinhetas iniciais e finais, ou intercalação de músicas ou de efeitos sonoros. A duração dos áudios processados e analisados oscilou entre 1m19s e 18min47s. O cálculo de  $f_0$  e da curva de intensidade foi realizado pelo software *Speech Filing System* (HUCKVALE et al., 1987; HUCKVALE, 2008) por autocorrelação. Posteriormente,



extraíram-se as matrizes de frequência e de intensidade. As matrizes foram analisadas pelo aplicativo ExProsodia (FERREIRA NETTO, 2010). As análises acústicas foram feitas com base na versão atual do *software* ExProsodia.

Os resultados foram interpretados com base nas hipóteses que estabelecem que o momento de maior tensão em narrativas tem como correlato acústico o aumento da frequência fundamental da curva entoacional (LABOV; WALETZKY, 1967; CHENOWETH, 1986; LONGACRE; CHENOWETH, 1986; LABOV, 1997; WENNERSTROM, 2001a). O parâmetro principal para a avaliação automática da entoação é decorrente da extração da média acumulada no tempo. O cálculo dessa média, a que chamamos de tom médio (*TM*), se faz mediante a aplicação da equação em (1).

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

Equação (1). Equação utilizada para a definição do Tom Médio e seus componentes.

Na equação (1), a variável *t* corresponde aos *n* instantes de uma elocução, medidos em saltos de 0,25 milissegundos. A transformação da curva de  $f_0$  em série temporal se dá pela obtenção dos momentos  $Z_{(t)}$ , cuja condição de existência é a duração mínima ( $t > 1ms$ ), intensidade mínima ( $I > 0 rms$ ) e do intervalo de frequência fundamental selecionado, desde que  $f_0 > 0$ . Os momentos  $Z_{(t)}$  foram por nós chamados de UBIs (unidades básicas da entoação) e são o segundo nível de análise do nosso modelo, sendo o primeiro a própria curva de *pitch*. O multiplicador ( $t - 1$ ) estabelece o número total dos momentos anteriores ao atual e atuam como pesos na média geral, (?) de modo a calcular a média acumulada no tempo. O Tom Médio é o conjunto da série de UBIs que se obteve com a decomposição apresentada em (1).

Uma vez definido o tom médio, estabelece-se uma faixa de 3 semitons acima e 3 semitons abaixo para cada um dos momentos  $Z_{(t)}$  (MARTINS, 2011), como se vê nas equações (2) e (3), chamados de limites laterais superior ( $l_{sup}$ ) e inferior ( $l_{inf}$ ). Esses valores correspondem à expectativa que se tem da faixa de frequência que será recuperada após a interrupção de produção sonora realizada no intervalo entre  $Z_{(t-1)}$  e  $Z_{(t)}$ .

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

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

O tom médio local (TMloc) é calculado após cada localização de final de frase entoacional (FERREIRA NETTO; MARTINS, 2018). A localização de final de frase entoacional se faz mediante a aplicação dos seguintes critérios, a serem usados alternativamente, mas não exclusivamente:

a) se entre dois momentos *Z* de  $F0$  ( $Z_{(t-1)}$  e  $Z_{(t)}$ ) consecutivos houver movimento descendente, por salto com intervalo maior do que o previsto e,

em seguida, movimento ascendente, por salto com intervalo maior do que o previsto, então  $Z_{(t-1)}$  é momento de finalização frasal e  $Z_{(t)}$  é momento inicial, portanto,  $t = 1$ .

b) se entre dois momentos  $Z$  de F0 ( $Z_{(t-1)}$  e  $Z_{(t)}$ ) consecutivos houver um movimento ascendente por salto (*shift*) com intervalo maior do que o previsto, sucedendo uma movimento descendente por salto cujo alvo estiver abaixo dos limites previstos para *F/Einf* (*TMloc* – intervalo previsto) e ainda anteceder uma pausa breve maior do que a prevista, então  $Z_{(t-1)}$  é momento de finalização frasal e  $Z_{(t)}$  é momento inicial, portanto,  $t = 1$ .

c) se entre dois momentos  $Z$  de F0 ( $Z_{(t-1)}$  e  $Z_{(t)}$ ) consecutivos houver uma pausa longa maior ou igual à 500 ms e um movimento ascendente por salto com intervalo maior do que o previsto ou um movimento descendente por salto com intervalo maior do que o previsto, então  $Z_{(t-1)}$  é momento de finalização frasal e  $Z_{(t)}$  é momento inicial, portanto,  $t = 1$ .

d) Se entre dois momentos  $Z$  de F0 ( $Z_{(t-1)}$  e  $Z_{(t)}$ ) consecutivos houver uma pausa longa maior do que duas vezes 500 ms, então  $Z_{(t-1)}$  é momento de finalização frasal e  $Z_{(t)}$  é momento inicial, portanto,  $t = 1$ .

Definidas essas fronteiras interfrasais, analisamos os seguintes critérios: movimento da entoação (movimento), direção da entoação (direção), alvo da entoação (alvo), interrupção do ritmo tonal (pausa >0,5 seg.), alongamento de UBI (alongamento), correlação entre tom médio e tom médio total (correlação), diferença entre o tom médio e o tom médio total (diferença) e dispersão de entoação acima do tom médio (F/E). A seguir apresentamos uma breve apreciação de cada um deles:

- Os critérios de *movimento*, *direção* e *alvo* da entoação relacionam-se com a variação de  $f_0$  em relação ao tom médio considerando seus intervalos laterais. Dessa maneira, entende-se que o movimento final de  $f_0$  pode ser *oblíquo* ou *pontual*; no caso de um movimento oblíquo, os valores finais de  $f_0$  serão próximos entre si, no caso de um movimento pontual, os valores finais de  $f_0$  serão distantes entre si, ocorrendo por salto entre UBIs contíguas. A direção de  $f_0$  pode ser *ascendente* ou *descendente*; as UBIs finais contíguas estarão em diferentes alturas, apresentando um movimento por salto ou oblíquo ascendente ou descendente. O alvo de  $f_0$  poderá ser *autêntico* ou *plagal*; no caso do alvo autêntico, espera-se que um movimento entoacional descendente que atinja um ponto grave, abaixo do limite inferior de  $f_0$ , calculado a partir da inequação (C); o alvo plagal caracteriza-se por manter a curva entoacional dentro dos limites de  $f_0$ , ou acima deles. Entende-se que a finalização por movimento descendente, pontual e autêntico seja a forma mais marcada da finalização de frases assertivas.
- A *interrupção do ritmo tonal* caracteriza-se mais comumente pela supressão intencional e demorada da fala do locutor. Tal supressão de fala pode ser causada tanto pela interferência de um interlocutor, em decorrência de troca ou de roubo de turno, como pela necessidade do próprio locutor. Nesta pesquisa, todas as pausas longas ( $\geq 0,5$  ms) foram consideradas fronteiras de frases entoacionais. Pressupõe-se que pausas longas sejam mais marcadas como fronteiras do que pausas com menor duração. Não foram conside-



- radas as “pausas preenchidas” nesta pesquisa.
- A *renovação* do tom médio, caracterizado pela presença de *shift* ou de *reset*, se estabeleceu pela diferença entre a frequência final de uma frase entoacional e a frequência inicial da frase seguinte. A renovação se definiu pela equação (A), em que há o reinício do momento  $Z_{(t)}$  do *TM* no multiplicador  $(t - 1)$ , bem como no do denominador  $(t)$ . Na medida em que a renovação interrompe o fluxo entoacional da fala, entende-se tratar-se de uma forma mais marcada do que sua inocorrência em frases assertivas.
  - Embora uma variação de duração segmental, por sua vez, não envolva qualquer relação com a variação entoacional da fala, foi tomada como um aspecto interno da formação das UBIs finais. Esse alongamento final se permite avaliar somente em comparação estatística (teste z) 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.
  - A *correlação* entre tom médio (*TM*) e tom médio local (*TMloc*), 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. Na medida em que o tom médio local se diferencia do tom médio pelo fato de que este não sofre interferências de fronteiras frasais e serve também como referência fixa para a avaliação do tom médio local, definido para cada frase entoacional. Assim, havendo uma declinação frasal bem definida ou muitas inflexões, a correlação entre tom médio e tom médio local será muito baixa ou inexistente. Uma correlação negativa também mostra diferenças marcadas entre eles. Dessa forma, assume-se que a ocorrência de forte correlação (próximo de +1) 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. Ao afastar-se da igualdade, que seria uma razão = 1, entende-se haver uma tendência maior à mudança de registro em relação ao que era esperado. 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.

Com o propósito de verificar se há regularidade na utilização desses critérios, o que apontaria a existência de regras para tanto, definimos a entropia de cada conjunto tomando a fórmula de entropia de Boltzman, adaptada por Shannon (1948). Assim, para cada conjunto de critérios

tomados, aditivamente, calculou-se  $H = p_{(x)} \times \log_2 p_{(x)}$  e, em seguida, a somatória de todos os valores encontrados para todas as frases entoacionais:

$H = -1 \sum p_{(x)} \times \log_2 p_{(x)}$ . Com o resultado  $H$ , fez-se sua razão com a entropia máxima definida para esse espaço amostral, mediante a fórmula

$H_{max} = \log_2 n/9$  e, então, calculou-se o que chamamos de Variação Relativa de Entropia ( $H_{vre}$ ), cuja fórmula é:

$$H_{vre} = \frac{H}{H_{max}} \#(4)$$

Entendemos que a variação relativa de entropia aponta para diferenças de restrições na seleção das alternativas disponíveis. Uma aproximação do valor da  $H_{vre}$  com o de uma entropia igual a zero ( $H_{zero}$ ) decorre de uma menor liberdade de escolha e/ou de uma maior restrição à seleção, independentemente do motivo que ocasionou essa restrição. Por se tratar de uma quantificação logarítmica, não linear, valores mais baixos tendem a aproximar-se mais rapidamente de zero e valores mais altos tendem a aproximar-

-se mais rapidamente da entropia máxima ( $H_{max}$ ). Assim, diferentemente da média aritmética, entendemos que o ideal será considerar a mediana como o ponto em que se define o centro das ocorrências.  $H_{zero}$  e  $H_{max}$

Assim, o estabelecimento de um valor mediano entre  $H_{zero}$  e  $H_{max}$  pode ser feito considerando as probabilidades mínimas de ocorrência de um espaço amostral com somente duas alternativas: 0-10; 1-9; 2-8; 3-7; 4-6; 5-5. As entropias desses sistemas são, respectivamente, 0, 0,467; 0,72; 0,886; 0,966 e 1. A mediana utilizada como referência — um valor de H entre 0,72 e 0,886 — define três campos específicos: abaixo de 0,72, entre 0,72 e 0,886 e acima de 0,886. O campo abaixo de 0,72 foi tratado como uma tendência ao estabelecimento de restrições fortes à ocorrência de algumas opções, com priorização de outras, portanto um processo de regulamentação de uso dos parâmetros. O campo acima de 0,89, como uma tendência ao abandono de restrições e, portanto, a um processo de aleatorização dos dados. Finalmente, o intervalo que incluiria os limites de 0,72 e 0,89, caracterizaria a homeostasia dos dados, isto é, o equilíbrio entre as tendências

máximas de estagnação ( $H_{zero}$ ) e de aleatorização ( $H_{max}$ ).

Tomando essa estatística como referência, foi possível verificar que a aplicação dos parâmetros propostos pelo projeto ExProsodia nas 10 narrativas populares analisadas disponíveis no Material Complementar, apontou para uma tendência a regularidade na medida em que seus valores de entropia relativa ocorreram todos abaixo do valor mínimo proposto de 0.72. O valor médio da entropia relativa foi  $H=0.6$  (máximo de 0.68 e mínimo de 0.47), com desvio padrão de 0.05. Esses valores apontam para a homogeneidade na aplicação dos recursos entoacionais previstos no projeto.

Discriminando os parâmetros, a entropia relativa de cada um apresentou variações bastante significativas.

**Tabela 1. Valores de entropia e entropia relativa de cada um dos parâmetros adotados para a segmentação frasal. Os valores marcados com cor vermelha ultrapassaram o limite superior do nível de homeostase referente à variação relativa de entropia**

	<i>H</i>	<i>H<sub>vrs</sub></i>
Direção	12,71389	0,847593
Movimento	14,03622	0,935748
Alvo	14,34606	0,956404
Alongamento	10,61442	0,707628
Correlação	13,60274	0,90685
Diferença	9,435061	0,629004
F/E	10,21059	0,680706
Pausa>0,5 seg.	13,61195	0,907463
Shift	13,17059	0,878039

Na tabela 1, podemos observar que os parâmetros de *movimento*, *alvo*, *correlação* e *pausa>0,5 seg.* ocorreram com variação com alto grau de aleatoriedade quanto à sua marcação, enquanto os demais — *direção*, *alongamento*, *diferença*, *F/E* e *shift* — apresentaram maior regularidade de ocorrência quanto a essa marcação.

No quadro 1 abaixo, fez-se a comparação de uma narrativa, utilizando de um lado as segmentações baseadas em unidades sentenciais definidas em Labov e Waletzky (1967) e Labov (1997) e de outro as segmentações automáticas baseadas no modelo ExProsodia. Na primeira e segunda colunas, as sentenças vão numeradas e classificadas respectivamente ação complicadora (ac), orientação (or) e juntura temporal (jt) conforme a proposição de Labov e Waletzky (1967) e Labov (1997). Na terceira e quartas colunas vão as descrições das sentenças e das frases entoacionais, numeradas e apresentadas em forma de gráfico, conforme a análise automática realizadas pelo aplicativo ExProsodia. A cor vermelha marca os textos em que o parâmetro *diferença* caracteriza o momento do evento mais relatável da narrativa. No Material Complementar desse artigo é possível verificar outras análises feitas utilizando o mesmo método.

**Quadro 1. Análise comparativa entre o método de Labov e análise automática da Rotina ExProsodia de uma narrativa do ciclo de Lampião. As siglas representam jt - juntura temporal, ac - sentença de ação complicadora e or - sentença de orientação. O sinal de positivo único + indica pausas com menos de 500ms, o sinal duplo ++ indica pausas com mais de 500ms.**

1	or	[1] Esse tempo de Lampião <u>foi foi</u> ++  [2] quase no mesmo tempo, <u>foi</u> vinte e seis isso, ++  [3] essa revolta. ++	
2	or	[4] Eu mesmo não <u>alcancei</u> ...	
3	or	Mas <u>via</u> ...	
4	or	<u>contava</u> ...	
5	or	Eu num... <u>dormiu</u> muito pro mato... ++	

6	or	[5] Inda <u>escutou</u> muita ah essas história...	
	jt	++	
7	or	<u>Escondia</u> cavalo dentro da rama no mato...	
	jt	++	
	jt	[6] Ai	
	jt	quando	
8	or	<u>chegava</u>	
9	or	o cavalo <u>relinchava</u> ...	
	jt	+	
10	or	ai se <u>assombrava</u> ...	
	jt	++	
11	ac	[7] Ele <u>ia</u> mais outro colega, quando	
12	ac	cavalo <u>relinchava</u> ...	
	jt	+	
13	ac	o cabra <u>ficava</u>	
	jt	+	
		[8] "Ave Maria [9] esse cavalo é um perigo!"	
14	ac	tudo <u>assombrado</u> com medo deles.	

### 3. Considerações finais

Os resultados obtidos pela análise da entropia da marcação dos parâmetros definidos, apontaram para uma seleção específica de padrões entoacionais nas narrativas que foram objeto desta pesquisa. O parâmetro referente à direção no final das frases entoacionais é um fato descrito formalmente desde a segunda metade do século XVIII (STEELE, 1779), mas foi na primeira metade do século XX que passou a ter um reconhecimento maior na descrição da língua (JONES, 1922; PIKE, 1945; TRUBETZKOY, 1973). Dentre os parâmetros analisados, foi o que apresentou maior ocorrência, tendo uma ocorrência média de 72%, nos dados. O parâmetro do alongamento, por sua vez, foi objeto de análise nos anos sessenta do século XX por Delattre (1966) que 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. Nas narrativas populares analisadas, entretanto, o alongamento só foi utilizado como marcador prosódico em 21% dos casos analisados, o que aponta para a sua baixa produtividade.

O parâmetro da diferença, que teve uma ocorrência igualmente baixa, sendo utilizado como marcador prosódico somente em 18% dos dados, foi objeto de discussão especialmente nas proposições que relacionavam a entoação com narrativas (LABOV; WALETZKY, 1967; CHENOWETH, 1986; LONGACRE; CHENOWETH, 1986; LABOV, 1997; WENNERSTROM, 2001a; b). O parâmetro de F/E, que marca frases entoacionais por meio de grandes variações do TM local, também teve ocorrência baixa, aparecendo em apenas 18% das frases. Em conjunto com o parâmetro da diferença, F/E caracteriza variações tonais acentuadas ao longo da frase em que ocorre, promovendo grande saliência local. Finalmente shift, que foi descrito na primeira

metade do século XX (PRONOVOST, 1939; FAIRBANKS, 1940; SNIDECOR, 1943; 1944; 1951), ocorreu em 67% das frases entoacionais. Os valores altos de entropia, por sua vez, que caracterizaram os parâmetros movimento, alvo, correlação e pausa > 0,5 seg, apontam para uma indiferença aparente do uso desses recursos na produção das narrativas que foram analisadas.

No âmbito do programa ExProsodia, o uso mais restrito ou mais abrangente de um parâmetro qualquer leva o locutor a evidenciar aspectos semânticos locais de sentenças em que as frases entoacionais ocorrem. Os valores baixos de entropia desses parâmetros corroboram a interpretação de que, nos textos que foram objeto desta pesquisa, apenas uma parcela se mostrou de acordo com o critério de sujeição a processos gramaticais específicos definido por Levinson (2007). Por se caracterizarem como parâmetros que se definem localmente nas frases entoacionais e não na narrativa como um todo, como se poderia verificar em variações do tom médio global (VASSOLER; MARTINS, 2013; FERREIRA NETTO et al., 2014; FERREIRA NETTO et al., 2017; MARTINS; FERREIRA NETTO, 2018), esses parâmetros relacionam-se diretamente com as intenções expressivas dos falantes quanto à produção de suas narrativas (FERREIRA NETTO, 2017; FERREIRA NETTO, 2018).

## AGRADECIMENTOS E CRITÉRIOS ÉTICOS

O material utilizado para a pesquisa está de acordo com a lei de direitos Lei Nº 9.610, de 19/02/1998, cap. IV, Art. 46, itens II e VIII. Em relação a ética na coleta e no uso dos dados, a coleta dos dados está conforme a Resolução do Plenário do Conselho Nacional de Saúde nº 510, 07/04/2016, Art 1º, parágrafo único itens V e VII. Além dos recursos da própria Universidade de São Paulo, esta pesquisa teve apoio do CNPq processos 400145/2009-0; PQ 300235/2010-0; 421369/2018-3.

## REFERÊNCIAS

ADOBE. **Adobe Audition 3.0.1 build 8347.0 [Audition\3.0]**. San Jose 2012.

BLOOMFIELD, L. **Language**. New York: Holt, Rinehart and Winston, 1961.

BOLINGER, D. L. **Intonation and Its Use: Melody in Grammar and Discourse**. Stanford: Stanford University Press, 1989.

CHENOWETH, V. Music as discourse. **Word-Journal of the International Linguistic Association**, v. 37, n. 1-2, p. 135-139, Apr-Aug 1986. ISSN 0043-7956.

CONCEIÇÃO, G. et al. Análise da ênfase prosódica em narrativas orais do ciclo de Lampião. **ExProsodia: resultados preliminares**. In: NETTO, W. F. (Org.) São Paulo: Paulistana Editora: 64-66 p. 2016.

CONSONI, F. **Aspectos da percepção da proeminência tonal em português brasileiro**. 2011. 119 (Doutorado). Departamento de Letras Clássicas

e Vernáculas, Universidade de São Paulo, São Paulo.

DARWIN, C. **A manifestação das emoções nos homens e nos animais**. São Paulo: Companhia das Letras, 2001.

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.

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

FERREIRA NETTO, W. **Variação de frequência e constituição da prosódia da língua portuguesa**. Tese de Livre-Docência. São Paulo: Universidade de São Paulo 2006.

FERREIRA NETTO, W. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. *In.*: CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA. **Anais[...]**. Niterói, 2008.

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

FERREIRA NETTO, W. **Tradição Oral, Narrativa e Sociedade**. São Paulo: Paulistana, 2017.

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.; 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.; MARTINS, M. V. M. Prosodic boundaries and automatic intonation analysis. *In.*: EICEFALA, 6. **Proceedings of the [...]**. Belo Horizonte: Universidade Federal de Minas Gerais, 2018. p.16

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

FREGE, G. Sense and Reference. **The Philosophical Review**, v. 57, n. 3, p. 209-230, 1948.

HUCKVALE, M. A. et al. The SPAR Speech Filing System. **European Conference On Speech Technology**. Edinburgh 1987.

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



JAKOBSON, R. Linguística e poética. In: BLIKSTEIN, I. (Ed.). **Linguística e comunicação**. 10. São Paulo: Cultrix, 1969. p.118-162.

JAKOBSON, R. Metalanguage as a linguistic problem. In: JAKOBSON, R. (Ed.). **The framework of language**. Michigan: University of Michigan, 1980. p.81-92.

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

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

LABOV, W.; WALETZKY, J. Narrative Analysis: oral versions of personal experience. In: HELM, J. (Ed.). **Essays on the Verbal and Visual Arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society**. Seattle: American Ethnological Society; University of Washington Press, 1967. p.12-44.

LEVINSON, S. C. **Pragmática**. São Paulo: Martins Fontes, 2007.

LONGACRE, R. E.; CHENOWETH, V. Discourse as music. **Word-Journal of the International Linguistic Association**, v. 37, n. 1-2, p. 125-134, Apr-Aug 1986.

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

MALINOWSKI, B. The problem of meaning in primitive languages. In: OGDEN, C. e RICHARDS, I. (Ed.). **The Meaning of Meaning . A study of the influence of language upon thought and of the science of symbolism**. New York: Brase & World, 1923. p. 296-336.

MARTINS, M. V. M. **Alvos tonais: unidades fonético-fonológicas da entoação**. 2017. 263 (Doutorado). Departamento de Letras Clássicas e Vernáculas, Universidade de São Paulo, São Paulo.

MARTINS, M. V. M. **Aspectos da percepção e do controle entoacional do Português Brasileiro**. 2011. 143 (Mestrado). Departamento de Linguística, Universidade de São Paulo, São Paulo.

MARTINS, M. V. M.; FERREIRA-NETTO, W. Prominence-marking mechanisms in emotional speech. **The Journal of the Acoustical Society of America**, v. 144, p. 1900, 2018.

MEAD, G. H. **Espírito, persona y sociedad. Desde el punto de vista del conductismo social**. Barcelona: Paidós, 1973.

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

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

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

RUSSEL, B. A filosofia do atomismo lógico. In: LACEY, H. M. (Ed.). **Ensaios Escolhidos**. São Paulo: Abril Cultural, 1978. (Os Pensadores).

SANTOS, A. J. dos. O Saci. In: NOGUEIRA, Leonardo; BONANOME, Vinícius. (eds.) **Memória Sanjoanense, fragmento da nossa história**. São João da Boa Vista: Prefeitura; 2008.

SAUSSURE, F. D. **Cours de Linguistique Générale**. Édition critique préparée par Tullio de Mauro. Paris: Payot, 1974.

SEARLE, J. R. **Consciência e Linguagem**. São Paulo: WMF Martins Fontes, 2010.

SEARLE, J. R. **Intencionalidade**. São Paulo: Martins Fontes, 2002.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, n. 3, p. 379-423, 1948

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.

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. 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.

STEELE, J. **Prosodia Rationalis: or an essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols**. Second edition amended and enlarged. London: J. Nichols, 1779.

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

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

VIDEOLAN. **VLC Mídia Player**. DENIS-COURMONT, R. e ALL., E. Boston:

Videolan Organization: Windows p. 2019.

WEBER, M. **Economia e Sociedade**. Fundamentos da sociologia compreensiva. Volume 1. Brasília: Editora Universidade de Brasília, 2015.

WENNERSTROM, A. Intonation an evaluation in oral narratives. **Journal of Pragmatics**, v. 33, p. 1183-1206, 2001a.

WENNERSTROM, A. **The music of everyday speech. Prosody and discourse analysis**. Oxford: Oxford University Press 2001b.