

8.01.01 – Teoria e Análise Linguística

O USO DA PROSÓDIA NO PROCESSAMENTO DA ESTRUTURA DO DISCURSOArthur R. B. Terto¹, Miguel Oliveira Júnior²

1. Estudante da Faculdade de Letras da Universidade Federal de Alagoas (FALE/UFAL)

2. Professor Dr. Associado da FALE – UFAL/Orientador

Resumo

O objetivo deste trabalho foi analisar as características de fronteiras prosódicas percebidas como proeminentes numa tarefa de segmentação de narrativas orais espontâneas em unidades comunicativas. Partindo da hipótese de que haveria um padrão acústico nessas fronteiras, analisamos diferença de F0, intensidade, duração e ocorrência de pausas silenciosas. Foram utilizadas trinta narrativas, e participaram do experimento cerca de 30 sujeitos. As análises estatísticas lançaram mão dos testes *modelo linear misto* e *regressão logística binomial*. Constatamos que as fronteiras percebidas como proeminentes pelos ouvintes possuem características prosódicas específicas: maior diferença de tom, maior diferença de intensidade e maior diferença de duração entre as sílabas tônicas que lhes são adjacentes. Esse trabalho fomenta o conhecimento dos efeitos da prosódia no processamento dos eventos no discurso falado.

Autorização legal: Número do parecer consubstanciado do CEP: 2.408.979

Palavras-chave: Eventos; Fronteiras prosódicas; Segmentação.

Apoio financeiro: CNPq.

Trabalho selecionado para a JNIC: Pró-reitoria de Pesquisa e Pós-Graduação da UFAL.

Introdução

Ao longo do dia, realizamos diversas ações: acordamos, tomamos café da manhã, vamos ao trabalho, almoçamos, saímos do trabalho, chegamos à nossa casa, jantamos e, por fim, dormimos. Percebemos as ações cotidianas como uma série de “pedaços” que podem ser identificados no tempo, os eventos (ZACKS et al., 2007). Eventos são ações (i) que tem um objetivo; (ii) que são situadas no tempo, ou seja, possuem um início e um fim; (iii) que ocorrem em um dado espaço; (iv) que envolvem pessoas e objetos etc. (RADVANSKY; ZACKS, 2014; ZACKS; TVERSKY, 2001).

Um número considerável de estudos tem investigado a relação entre a linguagem verbal e o processamento de eventos. Nesse contexto, há uma linha que se dedica a compreender de que modo funciona a representação de eventos na manifestação linguística de modalidade oral. Partindo do pressuposto de que os eventos podem ser representados na língua (ZACKS, 2020), surge este questionamento: que pistas da superfície linguística seriam tomadas pelo ouvinte para criar modelos de evento a partir de uma narrativa oral?

A literatura tem reportado que a prosódia é um fenômeno linguístico que serve para organizar as estruturas do discurso (BARBOSA, 2012). Esse papel da prosódia é visto também do ponto de vista da percepção. Os indivíduos parecem usar a prosódia para detectar e processar essas estruturas (SWERTS, 1996; SWERTS; GELUYKENS, 1994; OLIVEIRA JR., 2000; OLIVEIRA JR.; CRUZ; SILVA, 2012). Ela parece ser, portanto, uma pista importante para que o ouvinte perceba a estrutura de eventos no texto narrado. Resta saber, entretanto, que elementos prosódicos são mais salientes para o ouvinte na percepção das fronteiras de eventos, simbolizadas, em termos estruturais, pelas fronteiras prosódicas.

O objetivo geral desse estudo foi, portanto, investigar o papel dos elementos prosódicos no processamento da estrutura de eventos em narrativas orais espontâneas. Tal objetivo foi desdobrado em outros mais específicos, a saber: (i) coletar e segmentar dados de narrativas orais espontâneas; (ii) desenvolver experimento comportamental *on-line* em software livre para segmentação de narrativas espontâneas de língua portuguesa em unidades de evento; e (iii) identificar os elementos prosódicos que atuam nas fronteiras de eventos nas narrativas espontâneas.

Metodologia

Foram coletadas 10 narrativas de conversas espontâneas onde surgiam esporadicamente. Os critérios para a seleção das narrativas foram estes: (i) narrativas em que não houvesse sobreposição de vozes entre os interlocutores; (ii) narrativas com qualidade acústica, visando a uma maior acurácia na análise prosódica futura dos dados; e (iii) narrativas com, no máximo, 3 minutos e 30 segundos.

O processo de segmentação e revisão das narrativas se deu no ambiente do software *Praat*. Já o experimento comportamental foi desenvolvido na plataforma *on-line Gorilla Experiment Builder*. A tarefa do

experimento foi encaminhada por meio de um link para os e-mails e/ou outras ferramentas digitais dos participantes do experimento.

Participaram da pesquisa cerca de 30 sujeitos, com idades entre 18 e 30 anos. Os critérios para a seleção dos participantes foram: (i) ser falante de português como primeira língua; (ii) não ter problemas auditivos; (iii) não ter problemas neurológicos; (iv) não participar de pesquisas em fonética, mais especificamente em estudos sobre prosódia; e (v) ser estudante de graduação ou já ter concluído o ensino superior. Na tarefa, o participante tinha de ouvir as narrativas e segmentá-las, conforme sua percepção e experiência, em unidades comunicativas completas. O objetivo da tarefa era verificar, em um primeiro momento, se os participantes concordavam entre si nas segmentações feitas.

Para calcular a taxa de concordância entre os participantes, estabelecemos janelas para cada fronteira ao longo das narrativas cujos limites eram de 300 ms antes do fim da fronteira e 300 ms após a fronteira prosódica. Toda segmentação feita dentro desses intervalos foi computada para fins do cálculo de concordância entre os sujeitos. Consideramos que as fronteiras em que houvesse a marcação de, ao menos, 40% dos participantes, seriam categorizadas como fronteiras proeminentes. Durante essas análises e anotações, foram utilizados dois *scripts*: *BeatExtractor* e *Analyse Tier*.

Os elementos prosódicos analisados neste estudo estão relacionados a três principais parâmetros acústicos: a frequência fundamental (no caso da diferença de F0), a amplitude (no caso da intensidade) e a duração (no caso das pausas). Todos os valores referentes a esses três parâmetros foram extraídos, respectivamente, das curvas de F0 e de intensidade geradas automaticamente pelo *Praat* e de informações da extensão temporal dos períodos de silêncio que constituíam os trechos anotados no ambiente do software. Por meio desses recursos, os *scripts* supracitados coletaram informações que foram armazenadas em tabelas do *Excel* para posteriores análises estatísticas.

As análises estatísticas e o tratamento dos dados foram feitos no software *R* por meio destes pacotes: *readr*, *readxl*, *dplyr*, *ggplot2*, *lme4* e *lmerTest*. Os dois testes utilizados em nossas análises foram o modelo linear misto e a regressão logística binomial.

Resultados e Discussão

Em relação à diferença de F0, postulamos que aquelas fronteiras percebidas pelos ouvintes como proeminentes teriam uma maior diferença tonal entre o valor da frequência fundamental computada na sílaba tônica final da unidade prosódica finalizante e aquele computado na tônica inicial da unidade prosódica seguinte, pois, como estudos anteriores mostraram (PIJPER; SANDERMAN, 1994; OLIVEIRA JR., 2000, entre outros), é comum haver essa padronização quando fronteiras prosódicas coincidem com a percepção de mudança de eventos ou de tópico discursivo no texto narrado. Assim, ajustamos um modelo linear misto com *semitons* (*ref. 100 Hz*) como variável resposta e *posição de sílaba* (níveis: última sílaba, primeira sílaba), *proeminência* (níveis: não-proeminente, proeminente) e a interação entre essas variáveis como efeitos fixos. O modelo também continha interceptos aleatórios por janelas. Ao aplicarmos um teste anova para a comparação por modelos aninhados, verificamos que o modelo com interação entre as variáveis *posição de sílaba* e *proeminência* era significativo para explicar os dados ($p = 0.02$). Um teste *post-hoc* mostrou que não houve diferença de tom significativa entre as sílabas quando a fronteira era não-proeminente ($p = 0.39$). No entanto, quando a fronteira era proeminente, a diferença tonal entre as sílabas tônicas era significativamente maior ($p = 0.03$).

Em relação à intensidade, postulamos que aquelas fronteiras percebidas pelos ouvintes como proeminentes teriam uma diferença de intensidade significativa entre as unidades silábicas tônicas adjacentes a ela, uma vez que a queda do valor de intensidade, no fim de uma fronteira prosódica, e seu reinício, na unidade prosódica seguinte, também é um componente prosódico responsável pela sinalização de segmentações no discurso falado (ALMEIDA, 2017; BARBOSA, 2019, entre outros). Dito isso, ajustamos um modelo linear misto com *intensidade (dB)* como variável resposta e *posição de sílaba* (níveis: última sílaba, primeira sílaba), *proeminência* (níveis: não-proeminente, proeminente) e a interação entre estas variáveis como efeitos fixos. O modelo também continha interceptos aleatórios por janelas. Os contrastes entre os níveis das variáveis foram ajustados para o tipo *dummy coding*. Ao aplicarmos um teste anova para realizarmos uma comparação por modelos aninhados, verificamos que o modelo com interação entre as variáveis *posição de sílaba* e *proeminência* era significativo para explicar os dados de intensidade ($p < 0.001$). Por meio de um teste *post-hoc*, observamos que houve diferença significativa de intensidade entre as sílabas tônicas adjacentes às fronteiras tanto quando estas eram não-proeminentes ($p < 0.001$) quanto quando eram proeminentes ($p < 0.001$).

Em relação à duração das sílabas, postulamos que a diferença de duração entre a última sílaba tônica da unidade prosódica finalizante e a primeira sílaba tônica da unidade prosódica seguinte seria significativa, pois é comum que uma fronteira prosódica seja pré-anunciada por um alongamento da tônica final que a antecipa (BARBOSA, 2019). Nesse contexto, quanto mais importante a segmentação, em termos de sinalizar seções discursivas na fala, maior a saliência das fronteiras prosódicas (SWERTS; GELUYKENS, 1994). Assim, tal efeito seria maior, hipoteticamente, na condição em que a fronteira fosse percebida pelos ouvintes como proeminente, isto é, como uma fronteira prosódica que também correspondia a uma fronteira de eventos na estrutura discursiva. Ajustamos, portanto, um modelo linear misto com *duração (ms)* como variável resposta e *posição de sílaba* (níveis: última sílaba, primeira sílaba), *proeminência* (níveis: não-proeminente, proeminente) e a interação entre estas variáveis como efeitos fixos. O modelo também continha interceptos aleatórios por janelas. Os contrastes entre os níveis das variáveis foram ajustados para o tipo *dummy coding*. Ao rodarmos testes anova

para realizar uma comparação por modelos aninhados, verificamos que o modelo com interação entre as variáveis *posição de sílaba* e *proeminência* não era significativo para explicar os dados de duração ($p = 0.1281$), assim como também não o era o modelo que tinha apenas *proeminência* como efeito fixo ($p = 0.1895$). No entanto, um modelo com *posição de sílaba* como efeito fixo era significativo ($p < 0.001$). Os resultados mostraram, portanto, que, embora o fato de uma fronteira prosódica ser percebida como proeminente não explique a diferença de duração entre a última sílaba tônica de uma unidade prosódica e a primeira sílaba tônica da unidade prosódica posterior ($p = 0.18$), o vetor posição de sílaba é suficiente para explicar essa diferença de duração entre elas ($p < 0.001$).

Em relação à ocorrência de pausas silenciosas entre fronteiras, postulamos que haveria chances significativas de ocorrência de pausas silenciosas naquelas fronteiras que fossem percebidas como proeminentes, já que estudos anteriores observaram que as pausas são elementos cruciais na organização da estrutura do texto falado, sendo um dos elementos mais atuantes na sinalização das fronteiras discursivas (SWERTS, 1996; SWERTS; GELUYKENS, 1994). Para tanto, ajustamos uma regressão logística binomial com *pausa* (níveis: “Não”, significando não-ocorrência; “Sim”, significando ocorrência) como variável resposta e *proeminência* (níveis: NP, significando não-proeminente; P, significando proeminente) como efeito fixo. O contraste dos níveis da variável resposta (“Nao”, “Sim”) foi ajustado para *sum coding*, com “Nao” (que indica a não-ocorrência) como nível de referência da variável resposta. A regressão mostrou que as fronteiras percebidas como proeminentes têm chances muito significativas de conter uma pausa silenciosa ($p < 0.001$). No entanto, as chances de ocorrência de pausas também foram muito significativas quando as fronteiras não eram proeminentes ($p < 0.001$). Esses dados confirmam que a pausa é um elemento complementar e fundamental para a percepção de quaisquer fronteiras prosódicas.

Segundo a Teoria de Segmentação de Eventos (EST, na sigla em inglês) (RADVANSKY; ZACKS, 2014; ZACKS et al., 2007; ZACKS; TVERSKY, 2001), há cinco dimensões que desempenham um papel central na representação dos eventos a que somos expostos: tempo, espaço, entidade (uma personagem ou um objeto, por exemplo), causalidade e intencionalidade (motivo ou um objetivo específico) (ZWAAN, 2008). Havendo alterações em uma ou mais dessas dimensões, o indivíduo precisa reconstruir ou atualizar o modelo de eventos criado anteriormente, pois este já não comporta mais as informações que haviam sido integradas anteriormente (SPEER; ZACKS; REYNOLDS, 2007). Essas mudanças no contexto levam, geralmente, à percepção de uma fronteira de eventos – a finalização de um evento e o início de outro (ZACKS et al., 2007), as quais são sinalizadas, no caso de textos narrativos falados, tanto por informações linguísticas quanto paralinguísticas, dentre elas a prosódica (OLIVEIRA JR.; CRUZ; SILVA, 2012).

Em outras palavras, os resultados acima sugerem que a prosódia, independentemente de evidenciar um evento hierarquicamente maior ou menor (RADVANSKY; ZACKS, 2014), é usada não só na representação/percepção de eventos na superfície linguística, mas também na organização do texto falado como um todo. Ela parece antecipar ao ouvinte o fato de que haverá mudança em uma das cinco dimensões do modelo situacional de eventos (o que justifica a concordância entre os participantes na tarefa de segmentação), que será manifesta também, possivelmente, no nível segmental da fala. Por isso, sugerimos, de acordo com os resultados observados, que a representação dos eventos tem sustento, no que diz respeito à língua falada, nas informações suprasegmentais prosódicas.

Os resultados também sugerem, entretanto, que a sinalização das fronteiras de eventos na fala não é resultado de um único elemento prosódico, mas da atuação conjunta de todas as propriedades prosódicas analisadas. A atuação conjunta dos elementos prosódicos organiza a representação e a estruturação dos eventos ao longo da narrativa oral, o que explica o nível de concordância entre os sujeitos quanto ao processamento da estrutura do discurso e à segmentação deste em eventos.

Conclusões

A hipótese que guiou a análise feita neste trabalho sugeria que haveria um padrão nas fronteiras prosódicas percebidas como proeminentes pelos ouvintes. Pelos resultados observados, é possível dizer que a atuação conjunta dos elementos prosódicos analisados permite explicar a taxa de concordância por parte dos ouvintes na percepção da estrutura de eventos e na segmentação das narrativas orais espontâneas. Além disso, a diferença de F0 e a significativa presença de pausa silenciosa em todas as fronteiras prosódicas percebidas como fronteiras de eventos funcionam como marcadores de destaque na superfície linguística, representando uma confirmação de que determinado evento foi finalizado.

Sugerimos, para pesquisas futuras, o uso de técnicas on-line de processamento linguístico, como a pupilometria e a eletroencefalografia, pois os efeitos da percepção prosódica no processamento de eventos, em tempo real ainda não são conhecidos.

Referências bibliográficas

ALMEIDA, A. N. S. **Análise prosódica de agrupamentos numéricos no português do Brasil**. 2017. 400f. Tese (Doutoramento em Letras e Linguística) – Faculdade de Letras, Universidade Federal de Alagoas, Maceió, 2017.

- BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. **Rev. Est. Ling.**, Belo Horizonte, v. 20, n. 1, p. 11-27, jan./jun. 2012.
- BARBOSA, P. A. **Prosódia**. 1 ed. São Paulo: Parábola, 2019.
- OLIVEIRA, M., Jr. **Prosodic Features in Spontaneous Narratives**. Vancouver: Simon Fraser University, 2000.
- OLIVEIRA JR., M.; CRUZ, R.; SILVA, E. W. (2012). A relação entre a prosódia e a estrutura de narrativas espontâneas: um estudo perceptual. **Revista Diadorim / Revista de Estudos Linguísticos e Literários do Programa de Pós Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro**. Volume 12, 2013. <http://www.revistadiadorim.lettras.ufrj.br>.
- PIJPER, J. R.; SANDERMAN, A. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. **Journal of the Acoustical Society of America**, vol. 96, n. 4, p. 2037-2047, 1994. Disponível em: <https://doi.org/10.1121/1.410145>.
- RADVANSKY, G. A.; ZACKS, J. M. (2014). **Event cognition**. Oxford: Oxford University Press.
- SPEER, N.K.; ZACKS, J.M.; REYNOLDS, J.R. (2007). Human brain activity time-locked to narrative event boundaries. **Psychol. Sci.** 18, 449–455. <https://doi.org/10.1111/j.1467-9280.2007.01920.x>.
- SWERTS, M. (1996). Prosodic features at discourse boundaries of different strength. **The Journal of the Acoustical Society of America**, v. 101, n. 1, p. 514–521.
- SWERTS, M.; GELUYKENS, R. (1994). Prosody as a marker of information flow in spoken discourse. **Language and Speech**, v. 37, n. 1, p. 21–43.
- ZACKS, J. M. (2015). **Flicker: Your brain on movies**. (Oxford: Oxford University Press).
- ZACKS, J. M. **The Importance of Events in Conception and Language**. Leiden, The Netherlands: Brill, 2020.
- ZACKS, J. M.; SPEER, N. K.; REYNOLDS, J. R. (2009). Segmentation in reading and film comprehension. **Journal of Experimental Psychology. General**, 138,307–327. <https://doi.org/10.1037/a0015305>.
- ZACKS, J.M; SPEER, N.K; SWALLOW, K.M; BRAVER, T.S; REYNOLD, J.R. (2007). Event perception: A mind-brain perspective. **Psychol. Bull.** 133, 273–293.
- ZACKS, J. M.; TVERSKY, B. (2001). Event structure in perception and conception. **Psychol. Bull.** 127, 3–21. <https://doi.org/10.1037/0033-2909.127.1.3>.
- ZWANN, R.A., 2008. Time in language, situation models, and mental simulations. **Lang. Learn.** 58 (1), 13–26.