

O PAPEL DO LÉXICO NA VARIAÇÃO FONOLÓGICA: UMA ENTREVISTA COM THAÍS CRISTÓFARO-SILVA

Thaís Cristóforo-Silva¹

Camila De Bona²

thaiscristofarosilva@ufmg.br

camilabona@ifsul.edu.br

DE BONA – A redução de segmentos fonológicos tem desafiado tanto modelos de representações abstratas quanto modelos baseados em exemplares. Na sua opinião, quais são os principais alcances dessas abordagens e que desafios ainda precisam ser vencidos?

CRISTÓFARO-SILVA – Em primeiro lugar, eu gostaria de reavaliar a dicotomia entre “modelos de representações abstratas” e “modelos baseados em exemplares” indicada na primeira sentença da pergunta. Esta dicotomia parece indicar que exemplares não sejam construtos abstratos: e são! A questão em pauta, ao meu ver, diz respeito à natureza das representações. Vários modelos, geralmente rotulados como “formais”, sugerem uma representação **única** da qual são obtidas (por algum mecanismo formal) as representações reais de uso da fala. Por outro lado, na Teoria de Exemplares, a representação é compreendida como sendo **múltipla**, constituída dos vários exemplares registrados pela experiência real de uso da fala. É através de mecanismos de categorização que os exemplares se consolidam como categorias abstratas. Há

¹ Universidade Federal de Minas Gerais – UFMG. Pesquisadora do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

² Doutoranda; Universidade Federal do Rio Grande do Sul (UFRGS); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Instituto Federal de Educação, Ciência e Tecnologia Sul-riograndense (IFSul).

exemplares mais ou menos robustos dependendo da frequência em que foram atestados para fins de categorização. É a categorização que promove a emergência de categorias abstratas. Para discussão sobre a natureza das representações – simples ou complexas - eu recomendo a leitura do livro de Johnson e Mullenix (1997) e, em especial, do capítulo 1 que discute o mapeamento das representações mentais. Para a leitura em português sobre o tema, eu recomendo Cristóforo Silva (2003).

A aparente dicotomia apresentada na pergunta, de fato, reflete uma posição teórica. O debate não é quanto às representações serem abstratas ou não, e sim entre a representação ser simples (excluindo o detalhe fonético e parâmetros extralinguísticos) ou complexa (compreendendo o detalhe fonético e parâmetros extralinguísticos). As representações simples são tipicamente assumidas em modelos formais e as representações complexas são assumidas na Teoria de Exemplares.

Um segundo ponto da pergunta diz respeito aos alcances das abordagens formais e de exemplares. Tendo pressupostos diferentes, espera-se que os percursos a serem trilhados por cada modelo sejam diferentes. A elegância de modelos formais é encantadora! Modelos formais pautam-se na matemática discreta, ou finita, que é associada com a lógica formal e acomoda belamente a noção de contraste categórico e a organização lógica de unidades discretas. Por outro lado, a matemática contínua que é associada ao cálculo permite a análise de fenômenos contínuos e gradientes. A Teoria de Exemplares assume princípios da matemática contínua e a tendência atual é para modelos híbridos que consideram princípios da matemática contínua e discreta (Pierrehumbert, 2016). Para a discussão sobre a matemática discreta e contínua veja Cristóforo Silva (2006, p. 29-31).

A Teoria de Exemplares traz para a agenda de debate a interação entre a Fonética e a Fonologia. Modelos formais tendem a separar explicitamente o domínio destas disciplinas. Por outro lado, a Teoria de Exemplares acomoda a interação entre a Fonética e a Fonologia e busca explicar como estes domínios se articulam entre si (Cristóforo Silva & Gomes, 2017). A discussão, portanto, não é sobre o alcance das teorias, e sim sobre a natureza das categorias de análise: discretas ou gradientes?, e também quanto aos domínios de análise: Fonética ou Fonologia x Fonética e Fonologia.

Portanto, retomando a pergunta, o alcance da teoria tem a ver com suas premissas e seus objetos de análise. Por exemplo, um modelo que assuma categorias discretas em que a variabilidade é descartada não terá instrumentos óbvios para

analisar casos de variação linguística. Por outro lado, modelos que assumem categorias gradientes e representações múltiplas terá instrumentos claros para analisar casos de variação linguística. E, é na perspectiva da matemática contínua que efeitos de frequência têm sido considerados em linguística.

Retomando o início da pergunta – sobre a redução de segmentos fonológicos – o debate pauta-se na natureza do cancelamento: abrupto (de categorias discretas) ou gradiente (de categorias contínuas). Evidências robustas da literatura indicam que a redução segmental tende a ser gradiente, exceto em casos de analogia (cf. o debate lançado por Phillips (1984, 2001) e Bybee (2001)).

Finalmente, quanto aos desafios impostos para qualquer modelo, eu aponto a multimodalidade da fala que deve incorporar aspectos da produção, da percepção e outros aspectos relevantes para a comunicação oral ou gestual (visual, sensorial, etc). A modelagem multimodal lança grandes desafios metodológicos, mas, possivelmente, oferecerá explicações consistentes sobre a natureza da linguagem.

DE BONA – A relação entre variantes reduzidas e frequência lexical tem sido bastante discutida na literatura linguística recente. Em sua opinião, frequência lexical tem papel crucial (ou isolado) na redução segmental?

CRISTÓFARO-SILVA – A relação entre as diversas variantes reduzidas e não-reduzidas se dá ao longo de um contínuo gradiente. A concepção de que um segmento possa desaparecer abruptamente na fala é difícil de ser mantida. Obviamente, o ponto inicial e final do percurso de redução pode ser assumido como categoria discreta. Por exemplo, uma vogal é apagada em final de palavra: $V \rightarrow \emptyset / _ \#$. Sabemos pela literatura que o apagamento de vogais em final de palavra passa pelo desvozeamento de tais vogais (Dias & Seara, 2013). O desvozeamento de vogais seria um estágio gradiente na implementação do apagamento da vogal.

Portanto, se o foco de análise for categorias discretas, o pesquisador investigará os estágios inicial e final do fenômeno em questão. Por outro lado, se o foco de análise for a implementação do fenômeno - que eventualmente poderá ser concluído – a investigação será em categorias gradientes. Ou seja, o estudo de categorias gradientes só é foco de análise quando se investiga a implementação ou evolução de um fenômeno. Assim, surge uma questão: como fenômenos variáveis são implementados?

Uma hipótese de trabalho é que a implementação de mudanças sonoras ocorra inicialmente em palavras mais frequentes da língua. Tal hipótese foi consolidada em Bybee (2001, p.64) que sugere que “on-line adjustments that take place in production and have as their motivation the increased fluency of the sequences of gestures.”³ Ou seja, é a prática de padrões inovadores que reflete a redução (ou assimilação) de categorias. As formas reduzidas, inovadoras, se tornarão mais robustas e, eventualmente, a forma não-reduzida desaparece da língua. Uma vez que as palavras mais frequentes são usadas recorrentemente, temos que as formas inovadoras serão consolidadas nestas palavras. Obviamente, a frequência lexical não é o único fator que propulsiona uma mudança sonora. A frequência de tipo, parâmetros sociais, redes morfológicas, dentre outros fatores, podem contribuir para a implementação da mudança sonora em questão. Oliveira-Guimarães (2004) estudou a redução de [ʃtʃi] > [ʃi] em palavras como *triste* ou *plástico*. A autora observa que, de maneira geral, as palavras com o sufixo *-ik*, como por exemplo, *característica*, *estatística*, *artística*, favorecem a redução de [ʃtʃi] > [ʃi]. Embora haja o favorecimento da redução de [ʃtʃi] > [ʃi] em palavras com o sufixo *-ik*, a autora observa que palavras menos frequentes são afetadas mais tardiamente do que as mais frequentes: *cabalística*, *ritualístico*, etc. O estudo de caso discutido por Oliveira-Guimarães (2004) mostra que a redução de [ʃtʃi] > [ʃi] é motivada por sobreposição de gestos articulatórios afetando palavras mais frequentes inicialmente, mas favorecendo palavras que tenham o sufixo *-ik*.

Bybee (2001, p. 6) observa que “High-frequency words and phrases have stronger representations in the sense that they are more easily accessed and less likely to undergo analogical change.”⁴ Por esta razão, as mudanças de natureza gramatical ou analógica tendem a afetar as palavras menos frequentes primeiro. Temos, no português brasileiro, que um verbo pouco frequente como *pelejar* tende a apresentar pronúncias variáveis como *(ele/a) pel[e]ja* ou *(ele/a) pel[ɛ]ja* quando o prescrito seria *(ele/a) pel[e]ja*. Por outro lado, um verbo mais frequente como *desejar* não apresenta tal variação (Cristófaro Silva, T.; Campos, C. 2011). O estudo de caso apresentado em Cristófaro Silva e Campos (2011) oferece evidências para a frequência lexical em casos que não tenham motivação fonética.

³ Tradução livre da autora: “ajustes online que ocorrem na produção e têm como suas motivações o aumento da fluência de uma sequência de gestos (articulatórios).”

⁴ Tradução livre da autora: “Palavras e frases que tenham alta frequência têm representações mais fortes (robustas) no sentido que elas são mais facilmente acessadas e têm menor chance de se submeterem a mudanças analógicas.”

Resumindo, efeitos de frequência são relevantes na implementação de mudanças sonoras. Em casos de redução segmental e de assimilação, há evidências de que a evolução seja gradiente, podendo ou não haver a conclusão do fenômeno variável. As palavras mais frequentes são afetadas inicialmente. Por outro lado, em fenômenos gramaticais, as palavras menos frequentes são afetadas inicialmente e a mudança entre as categorias tende a ser abrupta.

DE BONA – Como você vê a relação entre frequência lexical e aspectos gramaticais, como morfologia interna à palavra ou categoria? E entre frequência lexical e fatores extralinguísticos?

CRISTÓFARO-SILVA – Em 1985, Joan Bybee escreveu um livro muito interessante chamado *Morphology*. Neste livro, a autora discute a natureza discreta ou gradiente das categorias morfológicas e pondera sobre efeitos de frequência na morfologia e na fonologia. Neste livro, a autora sugere um Modelo de Redes que para a época foi inovador e cuja proposta foi parcialmente incorporada na Teoria de Exemplos. Contudo, foi somente no final da década de 90 que efeitos de frequência passaram a ser discutidos de maneira abrangente. Em 2010, Joan Bybee escreveu o livro *Language and Cognition* que foi lançado em português em 2017. Neste livro, a autora apresenta uma abordagem de análise gramatical que considera efeitos de frequência em vários níveis. A análise apresentada em Bybee (2017) se beneficia da Gramática de Construções (Goldberg 1995, 2006; Rosario e Oliveira 2016; Oliveira 2013; Coelho & Silva 2014). Portanto, a relação entre frequência lexical e aspectos gramaticais decorre da concepção de Gramática assumida pelo pesquisador. Em uma gramática formal, com categorias discretas, em que o uso é excluído da análise não há qualquer justificativa para se considerar efeitos de frequência. O mesmo pode ser dito para a frequência lexical e os fatores extralinguísticos. A Teoria de Exemplos sugere que fatores linguísticos e extralinguísticos atuem nas representações mentais (Cristóforo Silva e Gomes 2017).

DE BONA – Nem sempre é simples definir uma escala ideal de frequência lexical ou decidir pelo melhor método estatístico de análise. Nesse aspecto, que recomendações você daria a um pesquisador nessa área?

CRISTÓFARO-SILVA – Há vários problemas em se mensurar frequência. O maior deles, possivelmente, é o de que seja impossível medir a frequência das palavras para uma comunidade, ou língua. Isso porque a frequência deveria ser medida para cada pessoa - uma vez que cada pessoa tem acesso diferente à frequência das palavras da língua. Contudo, a tarefa de medir a frequência lexical para cada falante de uma língua é inatingível.

Por outro lado, há indícios importantes de que a frequência lexical é um parâmetro importante na gramática das línguas. Por exemplo, verbos irregulares tendem a ser muito frequentes em várias línguas, como é o caso dos verbos *ir*, *ser*, *estar*, *comer*, etc. Este fato pode ser explicado por efeitos de frequência (Bybee 2001, p. 110). Devido ao uso frequente, estes verbos têm representações robustas e desenvolvimento diferente de verbos regulares. Por outro lado, verbos irregulares de baixa frequência ou caem em desuso ou se regularizam, como, por exemplo, os verbos *espelhar*, *velejar*, *ansiar*, *remediar* (Campos 2005). O trabalho clássico de Zipf (1929, 1949) e outros como Baker (1968) mostraram que a frequência das palavras tem organização particular: de maneira geral, há poucas palavras muito frequentes e inúmeras palavras pouco frequentes. Este padrão de organização de frequência é também encontrado em várias outras áreas da ciência como a Sociologia, Psicologia, Educação, etc.

Portanto, de um lado, é aparentemente impossível medir frequência lexical e, por outro lado, há evidências de que a frequência lexical seja relevante. A questão metodológica que se coloca é: o que é um corpus representativo para medir a frequência em uma língua? A ampla literatura em linguística de corpus tem contribuído significativamente para o desenvolvimento de corpora balanceados e representativos. Sugiro a leitura de Berber Sardinha (2004), Gries (2009) e McEnery & Hardie (2012) como introdução à linguística de corpus. Dentre os corpora para o português brasileiro, temos o NILC⁵, Corpus Brasileiro⁶, Portal Minas⁷, Corpus Eye⁸ e o C-Oral Brasil⁹. O Projeto Avaliação Sonora do Português Atual¹⁰ (ASPA) oferece busca de frequência com parâmetros fonológicos.

⁵ <http://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>

⁶ <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>

⁷ <http://portalminas.lettras.ufmg.br/>

⁸ <https://corp.hum.sdu.dk/cqp.pt.html>

⁹ <http://www.linguateca.pt/acesso/corpus.php?corpus=CORALBRASIL>

¹⁰ <http://www.projetoaspa.org/>

Respondendo objetivamente a pergunta formulada: Realmente não é simples se definir uma escala ideal de frequência lexical. Deve-se considerar a natureza contínua da frequência de um corpus e buscar os métodos adequados para o projeto que se pretende desenvolver.

Quanto às recomendações, eu sugiro que o pesquisador aceite os desafios impostos por uma análise de efeitos de frequência e busque modelos teóricos que acomodem tais efeitos. É a partir do poder explicativo das teorias, que conseguimos avançar na ciência. E, quanto aos métodos estatísticos, o importante é avaliar o conjunto de dados para que seja possível modelá-los de maneira apropriada. Os desafios são inúmeros! Mas, com as ferramentas adequadas, é possível vislumbrar um futuro muito interessante para o estudo da variação fonológica e sua contribuição para o conhecimento linguístico.

DE BONA – Que textos sobre o papel do léxico na variação fonológica você considera essenciais para sugerir a pesquisadores iniciantes na área?

CRISTÓFARO-SILVA – O debate sobre a perspectiva neogramática e da difusão lexical trouxe à baila a questão da frequência lexical na variação fonológica. Isto porque o debate entre as duas perspectivas centrou-se na questão de a mudança sonora se dar no **som** ou na **palavra** (Wang 1969, Labov 1981, Oliveira 1991). Tendo a palavra como objeto de análise na perspectiva difusionista, surgiu a questão da frequência lexical (Fidelholtz 1975; Phillips 1984, 2001). A partir deste debate, os trabalhos de Bybee (2001), Bybee & Hooper (2001), Pierrehumbert (2001, 2016) trouxeram um novo olhar sobre a natureza das representações mentais com ênfase na variação fonológica. Bybee (2017) é um texto importante em português sobre a natureza das representações mentais. Para efeitos de frequência na interação fonologia-morfologia, sugiro: Bybee (1988, 1995); Hay (2001), Losiewicz (1992). Alguns trabalhos em português que consideram o papel do léxico na fonologia: Albano et ali (1995), Gomes & Cristófaros Silva (2005), Cristófaros Silva (2006), Cristófaros Silva & Gomes (2007).

REFERÊNCIAS

- ALBANO, E. C.; MOREIRA, A. A. ; AQUINO, P. A. ; SILVA, A. H. P. ; KAKINOHANA, R. K. . Segment Frequency And Word Structure In Brazilian Portuguese. In: XIIIth International Congress of Phonetic Sciences, 1995, Estolcomo. Proceedings ICPHS 95. Estocolmo - Suecia, v. 3. p. 346-349. 1995.
- BAKER, J. Frequency in usage and in the lexicon. *Language*, v. 21, p. 13-22, 1968.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Manole. 2004.
- BYBEE, J. *Language Use and Cognition*. Cambridge: Cambridge University Press, 2010.
- BYBEE, J. *Língua, Uso e Cognição*. (tradução de *Language Use and Cognition*, 2010). Editora Cortez, 2017.
- BYBEE, J. Morphology as lexical organization. *Theoretical morphology*, ed. by M. Hammond and M. Noonan, 119–41. San Diego, CA: Academic Press. 1988.
- BYBEE, J. *Morphology: a study of the relation between meaning and form*. Philadelphia: Benjamins. 1985.
- BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.
- BYBEE, J. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–55. 1995.
- BYBEE, J. The phonology of the lexicon: evidence from lexical diffusion. *Usage-based models of language*, ed. by M. Barlow and S. Kemmer, 65–85. Stanford, CA: CSLI. 2000.
- BYBEE, J.; HOPPER, P. (ed.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001.
- CAMPOS, C. Abertura vocálica em verbos irregulares da primeira conjugação do português: um caso de reestruturação fonotática por generalização fonológica. Dissertação de Mestrado. FALE-UFMG. 2005.
- COELHO, S. M.; SILVA, S. E. P. . O continuum de gramaticalização do verbo DAR: de predicador a auxiliar. *SCRIPTA (PUCMG)*, v. 18, p. 23-40, 2014.
- CRISTÓFARO SILVA, T. ; CAMPOS, C, S. de. Gradiência fonética e generalização fonológica. *Leitura (UFAL)*, v. 46, p. 83-100, 2011.
- CRISTÓFARO SILVA, T. ; GOMES, C. A . Representações múltiplas e organização do componente lingüístico. *Fórum Linguístico (UFSC. Impreso)*, Florianópolis - Santa Catarina, v. 4, p. 147-177, 2007.
- CRISTÓFARO SILVA, T. Descartando fonemas: a representação mental na fonologia de uso. In: Dermeval da Hora; Gisella Collischonn. (Org.). *Teoria Lingüística: Fonologia e Outros Temas*. João Pessoa: .Universidade Federal da Paraíba, p. 200-231. 2003.
- CRISTÓFAROSILVA, T. Fonética e Fonologia: Perspectivas Complementares. *Revista de Estudos da Linguagem, Vitória da Conquista - BA*, v. 3, p. 25-40, 2006.
- CRISTÓFARO SILVA, T. Modelos Multirrepresentacionais em Fonologia. In: Renata Coelho Marchezan; Arnaldo Cortina (org). (Org.). *Os fatos da linguagem, esse conjunto*

- heteróclito. 1 ed. Araraquara: FCL-UNESP Laboratório Editorial. São Paulo: Cultura Acadêmica, p. 171-186. 2006.
- DIAS, E. C. O. ; SEARA, I. C. . Redução e apagamento de vogais átonas finais na fala de crianças e adultos de Florianópolis: uma análise acústica. *Letrônica*, v. 6, p. 71-93, 2013.
- FIDELHOLTZ, J. Word frequency and vowel reduction in English. *Chicago Linguistic Society* 11: 200–13. 1975.
- GOLDBERG, A. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press. 1995.
- GOLDBERG, A. Learning Linguistic Patterns. Categories in Use. In: MARKMAN, A.; ROSS, B. (eds.) *Psychology of Learning and Motivation*. 47. Academic Press/Elsevier. 2006.
- GOLDINGER, S. Words and voices: perception and production in an episodic lexicon. In: Johnson, K.; Mullenix J. *Talker variability in Speech Processing*. pp. 33-66. San Diego. Academic Press. 1997.
- GOMES, C. A.; CRISTÓFARO SILVA, T. Variação lingüística: antiga questão e novas perspectivas. *Lingua(gem)*, Amapa, v. 1, n.2, p. 31-41, 2005.
- GRIES, S. *What is corpus linguistics. Language and Linguistics Compass* 3 (2009): 1–17. 2009.
- HAY, J. Lexical frequency in morphology: is everything relative? *Linguistics* 39: 1041–70. 2001.
- JOHNSON, K.; MULLENIX J. Complex representations used in speech processing: overview of the book. In: Johnson, K.; Mullenix J. *Talker variability in Speech Processing*. pp. 1-8. San Diego. Academic Press. 1997.
- JOHNSON, K.; MULLENIX J. *Talker variability in Speech Processing*. Academic Press. San Diego. 1997.
- LABOV, W. Resolving the Neogrammarian controversy. *Language* 57.267–308. 1981.
- LOSIEWICZ, B. L. 1992. The effect of frequency on linguistic morphology. Dissertation. University of Texas, Austin, Texas.
- McENERY, T.; HARDIE, A. *Corpus Linguistics: Method, theory and practice*. Cambridge:Cambridge University Press. 2012.
- NOSOFSKY, R. M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: learning, memory, and cognition* 14: 54–65.
- OLIVEIRA GUIMARÃES, D. Sequências de (sibilante + africada alveopalatal) no português falado em Belo Horizonte. Dissertação de Mestrado. FAL-UFMG. 2004.
- OLIVEIRA, M. A. de. The Neogrammarian Controversy Revisited. *International Journal Of The Sociology Of Language, BERLIN - ALEMANHA*, v. 89, n.1, p. 93-105, 1991.
- OLIVEIRA, M. R. Gramaticalização de construções como tendência atual dos estudos funcionalistas. *Estudos Linguísticos (São Paulo)*, v. 42, p. 148-162, 2013.
- PHILLIPS, B. Lexical diffusion, lexical frequency, and lexical analysis. In: BYBEE, Joan; HOPPER, Paul. (ed.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001.

- PHILLIPS, B. Word frequency and the actuation of sound change. *Language*. 60.320–42. 1984.
- PIERREHUMBERT, J. Exemplar dynamics: word frequency, lenition and contrast. In: BYBEE, J.; HOPPER, P. (Ed.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, p. 137-157. 2001.
- PIERREHUMBERT, J. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*. 2, 33-52. 2016.
- RACZ P., PAPP V. AND HAY J. Frequency and Corpora. In Hippseley A; Stump G (Ed.), *The Cambridge Handbook of Morphology* Cambridge: Cambridge University Press. 2014.
- ROSARIO, I. C.; OLIVEIRA, M. R. . Funcionalismo e abordagem construcional da gramática. *Alfa: Revista de Linguística (UNESP. Online)*, v. 60, p. 233-259, 2016.
- WANG, W. S.-Y. Competing changes as a cause of residue. *Language*, 45:9-25. 1969.
- ZIPF, G. K. *Human Behavior and the Principle of Least Effort: an Introduction to Human Ecology*. New York: Hafner. 1949.
- ZIPF, G. K. *Relative frequency as a determinant of phonetic change*. *Harvard Studies in Classical Philology* 15: 1–95. 1929.