



TRUNCAMENTO NO PORTUGUÊS BRASILEIRO: UM ESTUDO DE FREQUÊNCIA COM DADOS DO *TWITTER*

TRUNCATION IN BRAZILIAN PORTUGUESE: A FREQUENCY STUDY WITH DATA FROM *TWITTER*

*Débora Heineck*¹

RESUMO

O truncamento é um processo não concatenativo de formação de palavras que, diferentemente dos processos de prefixação e sufixação, não se estrutura a partir da adjunção de afixos, mas sim a partir da supressão de segmentos da palavra-base, que não é, necessariamente, uma supressão de afixos. Por isso, o truncamento não se enquadra em análises formais de processos lineares de formação de palavras e foi considerado por alguns autores como um processo idiosincrático. Tendo como base teórica a proposta de Bybee (2004), esse estudo objetiva verificar a frequência de uso de dados de truncamento e analisar se esse fator exerce influência no processo de truncamento no português brasileiro. Para tanto, utilizaram-se dados de truncamento obtidos nos experimentos realizados em minha dissertação de Mestrado (HEINECK, 2018). A frequência desses dados foi verificada em uma busca realizada na rede social *Twitter*, utilizando um pacote do programa computacional R (*rtweet*). Os resultados da busca realizada no *Twitter* parecem confirmar algumas de nossas hipóteses: (a) os falantes conseguem recuperar a palavra-base mesmo de truncamentos frequentes na língua e (b) os dados mais frequentes nos experimentos realizados na Dissertação de Mestrado são também os mais frequentes na busca no *Twitter*, o que levanta a suposição de que os falantes já conhecessem essas formas. Por fim, a frequência dos diminutivos de *foto* e *moto* aponta para um resultado diferente ao encontrado nos experimentos, mas segue tendência mais geral encontrada para outros dados nos experimentos. Ao formar o diminutivo de uma forma truncada, os falantes parecem escolher entre as vogais finais *a* ou *o* de acordo com o gênero da palavra-base, independente da vogal final da forma truncada.

Palavras-chave: Truncamento; Processos não concatenativos; Morfofonologia; Frequência lexical.

ABSTRACT

Truncation is a non-concatenative word-formation process that, differently from prefixation and suffixation, is not structured by affix adjunction but instead by deletion of segments from the base. This deletion of segments on truncation process is not, necessarily, a deletion of affixes. Because of that, truncation does not fit in formal analyses of linear word-formation processes and it was considered by some authors as an idiosyncratic process. The theoretical base of this study is Bybee (2004) propose, and it aims to verify the frequency of use of truncation data and to analyze whether this factor influences the truncation process in Brazilian Portuguese. In order to do that, we used truncation data obtained in experiments carried out in my Master's Dissertation (HEINECK, 2018). The frequency of these data was verified in a search made on the social network *Twitter*, using a package of the computer program R (*rtweet*). The results of the search on *Twitter* seem to confirm some of our hypothesis: (a) speakers can recover the base-word even of frequent truncations and (b) the most frequent data in the Dissertation's experiments are also the most frequent data in the search on *Twitter*, which can point to the fact that speakers already knew these forms. Lastly, the frequency of *foto* and *moto*'s diminutive forms points to a different result from the one found in the experiments, but it follows the most general tendency found to other data in the experiments. To make the diminutive form of a truncation, speakers seem to choose between the final vowels *a* or *o* according to the gender of the base-word, regardless of the final vowel of the truncated form.

Keywords: Truncation; Non-concatenative processes; Morphophonology; Lexical frequency.

¹ Doutoranda em Estudos da Linguagem, PPG-LETRAS UFRGS (Departamento de Linguística, Filologia e Estudos Literários/Universidade Federal do Rio Grande do Sul), heineck.debora@gmail.com

Introdução

O português é uma língua predominantemente aglutinante ou concatenativa, ou seja, caracteriza-se pelo encadeamento de seus morfemas e pela formação de palavras através de processos lineares. Por esse motivo, de acordo com Gonçalves (2004), estudos sobre o português tendem a conceber a morfologia como “um módulo sintagmaticamente determinado pelo encadeamento de formativos” (p.10), e o português, de modo geral, se enquadra bem a esse tipo de estudo. Há, no entanto, alguns processos de formação de palavras que envolvem padrões não lineares de formação. Esses processos são denominados como não concatenativos pela ausência de encadeamento de elementos na sua formação.

O truncamento é considerado por diversos pesquisadores como um processo não concatenativo de formação de palavras que, diferentemente dos processos de prefixação e sufixação, não se estrutura a partir da adjunção de afixos, mas da supressão de segmentos da palavra-matriz (BELCHOR, 2014). Essa supressão de segmentos que ocorre no truncamento não é, necessariamente, uma supressão de afixos, como é possível ver nos exemplos em (1). Por esse motivo, o truncamento não se enquadra em análises formais de processos lineares de formação de palavras, e é caracterizado por autores como Gonçalves (2004) como um processo não concatenativo de formação de palavras.

- | | | |
|-----------------------|----------------------|------------------------|
| (1) delega < delegado | salafrá < salafrário | Maraca < Maracanã |
| sapa < sapatão | analfa < analfabeto | estranja < estrangeiro |
| cerva < cerveja | gurja < gorjeta | vagaba < vagabunda |

Fonte: Gonçalves (2004)

O truncamento, justamente por não se enquadrar na descrição de processos aglutinativos do português, foi considerado por alguns autores como um processo assistemático e idiossincrático. Nas gramáticas tradicionais, por exemplo, não há uma descrição sistemática do truncamento. Não há tampouco critérios para distinguir o truncamento de outros processos semelhantes a ele e, por isso, ele é geralmente descrito junto a processos como a siglagem, acronímia e abreviação².

As definições de Cunha e Cintra e Bechara se assemelham por usar do termo *abreviação* para denominar tanto o truncamento quanto a siglagem. Cunha e Cintra (1985) usam como

2 Alguns autores, como Gonçalves (2004) e Cristóforo Silva (2011) entendem os processos de siglagem, acronímia e abreviação como o mesmo processo. Em seu Dicionário de Fonética e Fonologia, Cristóforo Silva (2011, p. 201) define siglagem como “fenômeno que consiste na combinação de iniciais de sons que ocorrem em uma ou mais palavras também denominado redução sintagmática, acronímia ou abreviação”. Um exemplo deste fenômeno seria a sigla UFMG para Universidade Federal de Minas Gerais. Para alguns autores, estes são processos distintos. Para Carvalho (2010, apud LIMA, 2014, p. 39), siglagem se distingue da acronímia, pois, nas siglas, as letras conservam seu valor de soletração (PSB, por exemplo) e, nos acrônimos, consoantes e vogais tiradas da palavra abreviada formam uma nova palavra (*Ibope*, por exemplo). Rocha (1998) considera siglagem e acronímia como sinônimos, mas salienta que não devem ser confundidos com a abreviatura (ou abreviação), pois o resultado da abreviatura (ex.: *apartamento* > *apto*) não se constitui em uma nova palavra do português, pois não tem existência na linguagem oral, sendo um recurso apenas da linguagem escrita.

justificativa para a ocorrência do processo a tendência do mundo de hoje de “economizar tempo e palavras” e, por isso, fazer uso de uma “elocução mais rápida” (p. 114). Bechara (2006) afirma que o processo é comum tanto na fala coloquial, quanto na linguagem cuidada, por “breviedade de expressão” (p. 309). Ambos os autores, portanto, usam a economia linguística como justificativa para a ocorrência do processo que chamam de *abreviação*. No entanto, nenhum dos gramáticos dedicou-se a uma análise mais precisa e sistemática do processo.

Nos manuais de morfologia na perspectiva da ciência linguística, por outro lado, os autores fazem uma diferenciação mais acurada dos processos de encurtamento, separando truncamento de processos como acronímia e abreviação, por exemplo. No entanto, esses estudos ainda carecem de um olhar mais preciso para o funcionamento desses processos, visto que eles são ainda descritos como imprevisíveis.

Tanto nas gramáticas tradicionais quanto nos manuais de morfologia citados, portanto, o truncamento é visto como um processo idiossincrático e os autores não procuram investigar e descrever seu funcionamento. O truncamento, principalmente nas gramáticas tradicionais, é entendido como o mesmo fenômeno ou como um fenômeno semelhante a outros processos distintos da língua, como a siglagem, a abreviação e a acronímia. O truncamento é considerado um processo marginal da língua portuguesa e, por isso, não há uma preocupação em procurar regularidades no processo.

Autores como Basilio (1991), Sandmann (1992), Villalva (2007) admitem que há alguma regularidade em formas truncadas derivadas de compostos, como *foto* < *fotografia* e *mini* < *minissaia*, pois, nesses casos, de acordo com Villalva, “o truncamento parece reconhecer a estrutura de uma composição morfológica, preservando o primeiro radical (que é geralmente um radical neoclássico) e a vogal de ligação.” (VILLALVA, 2007, p.51). No entanto, em casos como *boteco* < *botequim*, *portuga* < *português* e *cerva* < *cerveja*, os autores não veem nenhuma regularidade e consideram o processo como arbitrário.

Nos estudos de Araújo (2002), Gonçalves (2004), Martini (2010), Scher (2011, 2016), Belchor (2009 e 2014), entre outros, o truncamento é analisado com o objetivo de encontrar alguma regularidade no processo. Para tanto, os autores baseiam-se em teorias morfológicas que lhes dão respaldo para explicar o funcionamento de processos como o truncamento. Grande parte desses autores fazem uso de teorias que aliam primitivos morfológicos e prosódicos, como é o caso da Morfologia Prosódica (MP) e da Teoria da Otimidade (TO) (e suas subteorias como, por exemplo, a Teoria da Correspondência). Esses autores, acreditam que o truncamento não é um processo puramente morfológico, mas sim, que tem acesso a primitivos prosódicos em sua formação. Dentre os autores mencionados, apenas Scher (2011 e 2016) realiza uma análise a partir da Morfologia Distribuída, na qual a fonologia é tardia e não interage com as operações morfológicas. Por esse motivo, a autora defende uma abordagem puramente morfológica do truncamento.

Em Heineck (2018), realizou-se uma investigação e descrição do processo do truncamento no português brasileiro. A partir da revisão da literatura, levantaram-se dados de truncamento

e esses dados foram organizados em um *corpus* de acordo com critérios morfológicos e fonológicos tanto da base (palavra-matriz) quanto do produto (forma truncada). Os dados do *corpus* serviram de base para a elaboração de dois experimentos. A análise dos dados bem como dos resultados dos experimentos teve como objetivos investigar o papel do acento e do número de sílabas da palavra-base na formação de truncamentos, analisar o comportamento de truncamentos formados a partir de bases compostas e com dois acentos fonológicos; e, por fim, investigar o *status* da vogal final do truncamento, verificando se ela se aproxima mais de uma vogal temática, de um morfema de gênero ou de um sufixo derivacional.

Neste estudo, em específico, iremos verificar a frequência de uso de dados de truncamento obtidos nos experimentos realizados em Heineck (2018). A busca é realizada na rede social *Twitter* e, para tanto, utiliza-se um pacote do programa computacional R – o pacote *rtweet*. A ferramenta permite tanto obter dados sobre a frequência de uso nesta rede social de certos itens lexicais quanto analisar o contexto no qual eles estão sendo empregados.

Na próxima seção deste trabalho, serão apresentadas teorias que tratam do papel da frequência lexical nos fenômenos linguísticos. Na sequência, a metodologia utilizada nos experimentos, na seleção e na análise dos dados será explanada. Por fim, apresentamos os resultados e as considerações finais da nossa análise.

Frequência lexical

Visto o objetivo deste trabalho de analisar se a frequência de uso exerce influência no processo de truncamento no português brasileiro, buscamos base teórica na proposta de Bybee, que introduz a ideia de que

a frequência com a qual palavras individuais ou sequências de palavras são usadas e a frequência com que certos padrões ocorrem em uma língua afetam a natureza das representações mentais e, em alguns casos, a forma fonética das palavras³ (BYBEE, 2004, p. 1, tradução nossa).

Bybee defende que o foco dado à estrutura em teorias anteriores, como o Estruturalismo, precisa ser complementado com uma perspectiva que inclua, além da estrutura, outros dois aspectos da língua: o conteúdo material (substância da linguagem) e o uso. A substância da linguagem está relacionada às duas extremidades – fonética e semântica – nas quais a língua é moldada e estruturada. O uso refere-se não apenas ao processamento da linguagem, mas a todos os usos sociais e interacionais da língua. Nesta proposta, substância e uso interagem e criam estrutura.

A teoria proposta por Bybee é conhecida por *Usage-Based Model* (traduzido como Fonologia de Uso em trabalhos brasileiros). A teoria é composta pelos princípios básicos que seguem (cf. BYBEE, 2004, p. 6-8):

3 “(...) the frequency with which individual words or sequences of words are used and the frequency with which certain patterns recur in a language affects the nature of mental representation and in some cases the actual phonetic shape of words.”

1. A experiência afeta a representação. O uso de formas e padrões afeta suas representações na memória. Palavras muito frequentes tem uma representação mais forte no sentido de que são mais facilmente acessadas e menos suscetíveis a mudanças analógicas.
2. Representações mentais de objetos linguísticos têm as mesmas propriedades de representações mentais de outros objetos. Isso implica que representações mentais não tem propriedades preditivas abstratas, mas que são baseadas em categorizações de itens reais.
3. Categorização é baseada em identidade ou similaridade.
4. Generalizações sobre formas não são separadas de representações de formas armazenadas, mas emergem diretamente delas.
5. A organização lexical fornece generalizações e segmentação em vários graus de abstração e generalização. Unidades como morfema, segmento ou sílaba são emergentes no sentido que eles surgem de relações de identidade e similaridade que organizam representações.
6. O conhecimento gramatical é conhecimento processual. Parte do conhecimento linguístico é declarativo (podemos citar o significado de palavras, por exemplo), mas grande parte do conhecimento linguístico é processual. Isso implica dizer que a fonologia se torna parte do processo de produzir e decodificar construções, e não mais um sistema psicológico e puramente abstrato.

Bybee (2004) propõe ainda que o conceito de frequência seja dividido em duas medidas diferentes: a **frequência de ocorrência** (*token frequency*) e a **frequência de tipo** (*type frequency*). A frequência de ocorrência indica quantas vezes uma determinada palavra ocorre em um *corpus* linguístico. A frequência de tipo indica quantos itens da língua contém uma determinada estrutura. Huback (2013, p. 83-84) mostra as consequências cognitivas e linguísticas da interação dessas duas medidas de frequência:

- palavras com alta frequência de ocorrência têm memória mais ativa e podem ser acessadas de forma independente no léxico mental; palavras com baixa frequência de ocorrência dependem de conexões com outros itens lexicais para ser ativadas;
- Palavras irregulares e frequentes têm memória mais forte no léxico mental, portanto costumam resistir a mudanças analógicas; por outro lado, palavras irregulares e infrequentes não são suficientemente reforçadas no léxico mental para manter sua irregularidade, então podem adotar os paradigmas mais frequentes da língua;
- Palavras derivadas e de frequência de ocorrência baixa ou média formam conexões mais fortes do que palavras derivadas e de alta frequência de ocorrência. Isso ocorre porque os itens infrequentes dependem de sua classe para ser lembrados, portanto, cada vez que são acessados, reforçam a coesão de sua rede; por outro lado, itens derivados altamente frequentes ganham autonomia lexical e sua ativação no léxico mental não é feita através de sua classe. Isso mostra a importância da frequência de tipo, não a de ocorrência, para a produtividade de uma classe.

Diversos trabalhos sobre variação e mudança linguística têm considerado a frequência lexical como uma de suas variáveis e vêm atestando que palavras de diferentes faixas de frequência podem ser afetadas de forma distinta pela mudança linguística. Huback (2013) traça uma linha do tempo dos estudos de mudança linguística que trataram de efeitos de frequência.

Dentre os trabalhos citados por Huback, está Fidelholtz (1975), que analisa a redução vocálica do inglês e observa que palavras mais frequentes estão mais suscetíveis ao fenômeno. Outro ponto importante atestado pelo autor é o papel da familiaridade da palavra, que, segundo ele, também determina se ela será afetada por fenômenos linguísticos. O autor apresenta o exemplo do topônimo “Manhattan” que seria mais reduzido por falantes da própria língua. Huback alerta para a importância de distinguir os conceitos de familiaridade e frequência. De acordo com a autora, “a familiaridade de um item nem sempre significa que ele será mais frequente na língua ou em um *corpus* específico” (HUBACK, 2013, p. 84). Objetos domésticos, como *mesa* e *cadeira*, por exemplo, são muito familiares para os falantes, mas a frequência desses itens provavelmente é baixa. Os conceitos de familiaridade e frequência, portanto, interagem, mas não são sinônimos.

Esta pesquisa pretende investigar se a frequência tem algum papel no processo de truncamento. Ao verificar a frequência de ocorrência de formas truncadas, objetivamos investigar se os falantes ainda conseguem recuperar a palavra-base de truncamentos muito frequentes na língua e se os dados mais frequentes nos resultados de Heineck (2018) são também os mais frequentes no *Twitter*. Na próxima seção, serão discutidos os procedimentos metodológicos que serão adotados de modo a analisar o papel da frequência lexical na tentativa de responder a essas perguntas.

Metodologia

Os dados selecionados para a análise de frequência são provenientes de dois experimentos realizados em estudo anterior (HEINECK, 2018). Por esse motivo, é necessário tratar da metodologia utilizada na construção e aplicação destes experimentos.

Primeiramente, antes da elaboração dos experimentos, realizamos um levantamento de dados de truncamento para a construção de um *corpus*. Parte dos dados do *corpus* foi extraída de trabalhos anteriores sobre esse processo (ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014). Além disso, ao longo do trabalho de levantamento, dados encontrados em redes sociais ou produzidos em situações de interação entre falantes foram registrados e também compuseram o *corpus*.

Os dados foram organizados de acordo com critérios fonológicos e morfológicos tanto da base submetida ao processo quanto de seu produto, isto é, da forma efetivamente truncada. Os critérios utilizados foram número de sílabas e padrão acentual do truncamento e de sua palavra-base, analisamos ainda se a forma truncada resultava de uma palavra composta por dois acentos fonológicos, como em *pós-graduação*, ou ainda se a palavra-base era composta de duas palavras morfológicas, com raízes eruditas, como, por exemplo, em *odontologia*.

Os dados do *corpus* foram utilizados em um experimento *on-line* e um experimento presencial. Ambos eram compostos de um questionário social e três exercícios. O experimento *on-line* foi ancorado na plataforma SurveyMonkey – disponível em <https://www.surveymonkey.com/> –, que permite a criação de questionários *on-line* para pesquisas de diversos fins. As respostas do nosso questionário foram completamente anônimas, o nome do informante não foi solicitado em nenhum momento e não tivemos acesso ao endereço de IP do respondente. A divulgação do experimento foi feita através de redes sociais e contou com 43 participantes.

O experimento presencial foi realizado em uma escola de Ensino Médio da rede privada de Porto Alegre com estudantes do primeiro e do segundo ano. Diferentemente do experimento *on-line*, esse experimento foi impresso em papel e os participantes respondiam às questões à mão. O experimento presencial contou com 46 participantes.

A primeira etapa do experimento consistiu em um questionário social. Os informantes informaram seu gênero, idade, escolaridade, curso superior (caso o informante optasse por curso superior completo ou incompleto no campo escolaridade) e local de residência (estado e cidade).

No experimento *online*, a grande maioria dos informantes era composta por mulheres: 37 dos 43% informantes (86%) e 81,2% dos informantes tinham entre 18 e 35 anos. No que diz respeito à escolaridade, todos os informantes possuíam Ensino Superior completo e incompleto e a grande maioria deles cursava ou havia cursado Letras (82,2%).

No experimento presencial, houve um equilíbrio maior entre o número de informantes homens e mulheres: dos 46 informantes, 26 eram mulheres (56,5%) e 20 eram homens (43,5%). A faixa etária dos informantes foi de 15 a 18 anos, sendo que metade tinha 15 anos. Todos eram estudantes do Ensino Médio.

Os dados do questionário social não foram considerados na análise desta etapa da pesquisa, que se propôs a uma análise ainda inicial e descritiva de dados de frequência do fenômeno do truncamento. Em futuras etapas, objetiva-se incluir esses dados na análise de modo a verificar sua relevância para o fenômeno.

Na segunda etapa do experimento, os informantes foram convidados a realizar três exercícios. Esses exercícios tinham, como objetivo geral, observar padrões no processo do truncamento.

No primeiro exercício, os informantes foram apresentados a uma lista de 30 formas truncadas e deveriam identificar a palavra-base dessas formas, como vemos na Figura 1.

Figura 1 – Exemplo do exercício 1 do experimento presencial

PERGUNTA 1

Abaixo, há uma lista de palavras encurtadas. Você deve preencher as lacunas com a palavra que você considera dar origem a cada uma dessas formas. As sílabas estão sublinhadas (como em visu) para indicar a sílaba mais forte (tônica) da palavra.

- visu _____
- burga _____
- condi _____
- japa _____

Com esse exercício, tínhamos como objetivo verificar se os informantes seriam capazes de identificar a palavra de origem das formas truncadas, se haveria concordância na escolha dessas formas e se elas seriam as mesmas do nosso *corpus*.

No segundo exercício do teste, foram apresentadas palavras do português (30 palavras) e os informantes deveriam propor uma forma truncada para cada palavra, como vemos na Figura 2. No experimento *on-line*, as palavras foram apresentadas em forma de lista, isoladas. Já no experimento *presencial*, as palavras estavam inseridas em sentenças, de modo que ficasse mais claro o sentido das palavras que estávamos apresentando.

Figura 2 – Exemplo do exercício 2 do experimento presencial

PERGUNTA 2

A partir das palavras **em destaque** abaixo, sugira uma forma encurtada da palavra que lhe pareça soar bem na fala informal. Sublinhe ou circule a sílaba ou vogal que você considera forte (tônica) em cada uma das palavras que você sugerir.

- A pesquisa tratava da análise da **macromolécula** de uma proteína.

- O **chinês** inaugurou uma nova loja.

- Não custou nenhum **tostão** para consertar o aparelho.

Nesse exercício, tínhamos como objetivo observar qual a estratégia usada pelos informantes para reduzir as palavras. Pretendíamos checar se o acento da palavra-base seria mantido ou se haveria algum padrão acentual preferido nas formas truncadas que funcionaria independentemente do acento da palavra-base e, ainda, se haveria sempre uma diminuição em número de sílabas da palavra-base em relação a sua forma truncada.

No terceiro e último exercício, os informantes foram apresentados a 20 formas truncadas e deveriam propor um diminutivo (com inho/inha/zinho/zinha) para essas formas, como vemos na Figura 3. Assim como ocorreu no exercício anterior, no experimento *on-line*, havia apenas uma lista de palavras, no entanto, no experimento presencial, os truncamentos estavam inseridos em sentenças para melhor compreensão do seu significado pelos informantes.

Figura 3 – Exemplo do exercício 3 do experimento presencial

PERGUNTA 3

Proponha um diminutivo para as palavras **em destaque** abaixo usando uma dessas terminações inho/inha/zinho/zinha.

- Vamos tomar uma **ceva**?
-

- O meu **computa** estragou de novo!
-

- Ela fez uma **lipo**.
-

Com esse exercício, tínhamos a pretensão de colaborar para a discussão sobre o *status* da vogal final do truncamento. O objetivo era verificar se, ao propor um diminutivo para a forma truncada, os informantes recuperariam o gênero da palavra-base ou se, ao invés disso, utilizariam a vogal final do truncamento. Pretendíamos verificar, portanto, se em um truncamento como *foto*, por exemplo, os informantes optariam pelo diminutivo *fotinho*, utilizando a vogal final do truncamento, ou *fotinha*, recuperando o expoente de feminino.

Os dados de truncamento analisados neste estudo foram dados retirados dos resultados dos dois experimentos (*on-line* e presencial). Definimos alguns critérios para selecionar os dados a serem buscados no *Twitter* para o estudo de frequência. No exercício 1, em que os participantes deveriam identificar a palavra-base de formas truncadas, foram selecionados apenas dados que tivessem 70% ou mais de concordância na escolha da palavra-base. Para o truncamento *japa*, por exemplo, 96% dos informantes apontou *japonês* como palavra-base, por isso, esse dado compôs a lista das formas truncadas que foram buscadas no *Twitter*. O objetivo é verificar se esses truncamentos em que a base é facilmente recuperável pelos falantes são bastante frequentes na língua e, dessa forma, formular hipóteses sobre o grau de transparência do processo de truncamento: será que os falantes ainda conseguem recuperar a palavra-base de truncamentos muito frequentes na língua?

Para a seleção dos dados do exercício 2, em que os informantes propunham formas truncadas para palavras do português, selecionamos dados com 60% ou mais de ocorrência nos resultados dos experimentos. O objetivo, aqui, é verificar se os truncamentos mais frequentes nos resultados são também frequentes no *Twitter*.

Por fim, realizamos a busca e verificamos a frequência de ocorrência no *Twitter* dos diminutivos de *foto* e *moto* – *fotinho*, *fotinha*, *motinho* e *motinha* – dados do exercício 3. O objetivo é contribuir para a discussão acerca do *status* da vogal final do truncamento.

Realizou-se, então, uma busca no *Twitter* dos dados selecionados. O *Twitter* é uma espécie de ‘microblog’, com um limite de 280 caracteres para cada publicação. Criado em 2006, hoje conta com cerca de 330 milhões de usuários em todo o mundo. Com base nessas informações, o *Twitter* parece ser um ambiente propício para a ocorrência de fenômenos como o truncamento, visto que se trata de um ambiente no qual o uso da linguagem é mais informal e, por seu limite de caracteres, talvez possa propiciar o encurtamento de palavras.

Para realizar a busca de dados no *Twitter* utilizou-se um pacote do programa computacional R – o pacote *rtweet*. Com esta ferramenta, é possível sejam feitas pesquisas no *Twitter* a partir de alguma palavra ou frase. A ferramenta permite controlar a localização e idioma dos *tweets* a serem buscados, bem como a inclusão ou exclusão dos chamados *retweets*, que são o compartilhamento de *tweets* já existentes. É possível, ainda, controlar a quantidade de dados que se deseja coletar na pesquisa. A ferramenta, no entanto, tem também algumas limitações. É possível buscar, no máximo, 18.000 *tweets* e os dados serão de até sete dias antes da data da busca.

Tendo em vista as possibilidades e limitações da ferramenta, definimos o desenho de nossa busca. A busca foi limitada a *tweets* publicados no Brasil e em língua portuguesa. Excluímos dados provenientes de *retweets*, para garantir que algum *tweet* que viralizou, ou seja, que tenha sido compartilhado muitas vezes, acabe não aumentando a frequência de algum dado. Realizaram-se três buscas com um intervalo de tempo de, no mínimo, uma semana entre elas, de modo a verificar se não haveria uma discrepância na frequência de ocorrência de algum dado nas buscas.

Nesta etapa do estudo, realizamos uma análise descritiva e quantitativa dos dados coletados no *Twitter*. A quantificação e análise dos dados foi realizada no programa *Excel*.

O resultado das buscas bem como a sua análise e discussão será feita na próxima seção deste trabalho. Os dados coletados no *Twitter* serão analisados juntamente com os resultados obtidos no estudo realizado em Heineck (2018), de modo a verificar o papel da frequência lexical no fenômeno do truncamento.

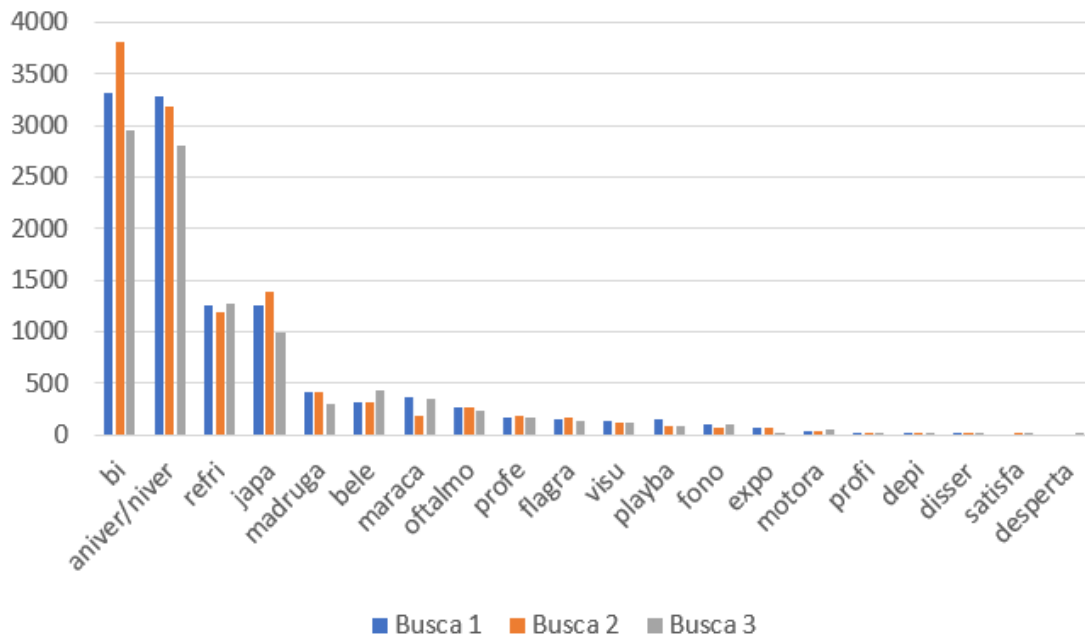
Resultados e Discussão

Devido às limitações da ferramenta utilizada, que permite buscar dados de, no máximo, sete dias antes da data da busca, os dados de frequência analisados neste trabalho são das últimas três buscas realizadas. A frequência de ocorrência dos dados no *Twitter* se mostrou estável, de um modo geral, nas três coletas realizadas, como vemos nos gráficos 1, 2 e 3. Isso confere maior confiança aos nossos dados, pois nos mostra que a frequência alta de algum dado,

por exemplo, não está ocorrendo em função de aquele dado estar relacionado a algum assunto muito discutido no momento.

O Gráfico 1 mostra a frequência de ocorrência no *Twitter* dos dados do exercício 1 nas três buscas realizadas. Como vemos no gráfico, de um modo geral, a frequência dos dados se mantém estável nas três buscas realizadas.

Gráfico 1 – Frequência de ocorrência dos dados do exercício 1 nas três buscas realizadas no *Twitter*

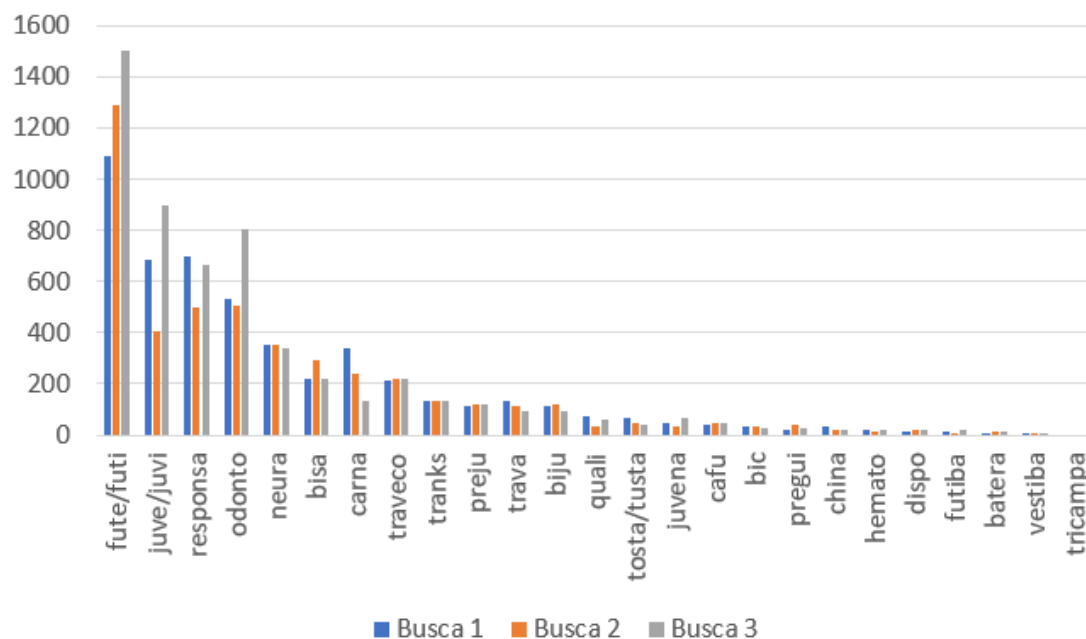


De modo a verificar se alguma das diferenças observadas no gráfico, como em *bi*, por exemplo, eram relevantes, calculamos a média da frequência de ocorrência das três buscas e verificamos a diferença absoluta de cada uma das buscas em relação a essa média. Excluímos desta análise os casos em que as buscas resultaram em pouquíssimos dados, em que a média não passou de 10 dados. Feita esta exclusão, a variação em relação à média não ultrapassou 20% na grande maioria dos dados. Os únicos três casos em que houve uma variação maior do que 20% são os representados na tabela abaixo, na qual apresentamos o número bruto de dados encontrados em cada busca, a média das três buscas e, entre parênteses, está a variação em relação à média.

Tabela 1 – Dados do exercício 1 com variação de frequência de ocorrência maior que 20% em relação à média da frequência nas três buscas realizadas no *Twitter*

Truncamento	Busca 1	Busca 2	Busca 3	Média das buscas
<i>maraca</i>	361 (20,07%)	189 (37,14%)	352 (17,07%)	300,67
<i>playba</i>	150 (41,96%)	80 (24,29%)	87 (17,67%)	105,67
<i>expo</i>	73 (37,74%)	67 (26,42%)	19 (64,15%)	53

O segundo gráfico representa a frequência de ocorrência dos dados do exercício 2 nas três buscas realizadas. Assim como no exercício 1, o gráfico parece mostrar que, de modo geral, há estabilidade nos dados das três buscas.

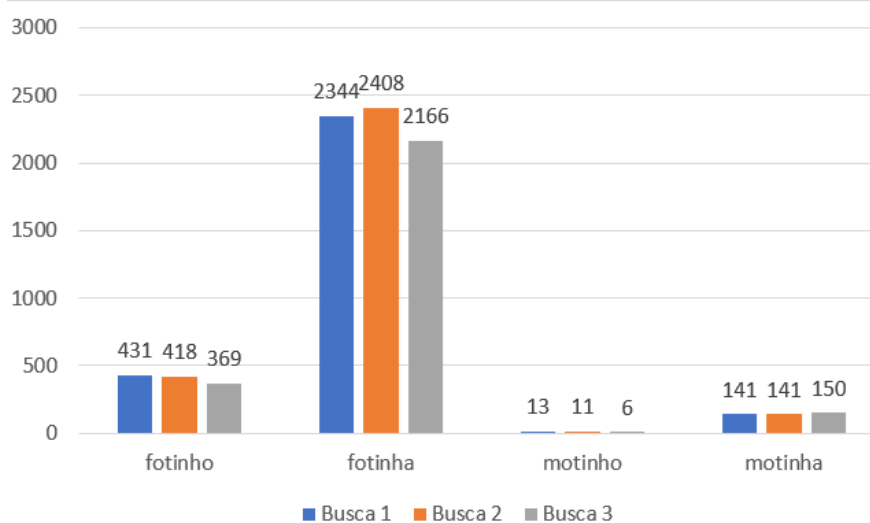
Gráfico 2 – Frequência de ocorrência dos dados do exercício 2 nas três buscas realizadas no *Twitter*

A análise de variação em relação à média parece confirmar a estabilidade dos dados verificada neste gráfico. Na maior parte dos dados, essa variação chegou, no máximo, em torno dos 30%. Na tabela abaixo, estão os 3 casos que apresentaram maior variação.

Tabela 2 – Dados do exercício 2 com variação de frequência de ocorrência maior que 30% em relação a média da frequência nas três buscas realizadas no *Twitter*

Truncamento	Busca 1	Busca 2	Busca 3	Média das buscas
<i>juve/juvi</i>	686 (3,57%)	406 (38,7%)	895 (35,13%)	662,33
<i>carna</i>	337 (41,8%)	241 (1,4%)	135 (43,2%)	237,67
<i>china</i>	35 (40%)	22 (12%)	18 (28%)	25,00

Por fim, o Gráfico 3 apresenta a frequência de ocorrência dos diminutivos dos truncamentos *foto* e *moto* nas três buscas. Vemos pelos dados brutos, que não houve grande variação no número de dados encontrados nas buscas.

Gráfico 3 – Frequência de ocorrência dos dados do exercício 3 nas três buscas realizadas no *Twitter*

Nas subseções a seguir, será realizada uma análise mais cuidadosa dos dados de frequência dos truncamentos de cada um dos exercícios. Essa análise visa responder as questões norteadoras deste estudo.

Dados do exercício 1

Conforme explicitado na seção de Metodologia, no exercício 1 de nossos experimentos, ao serem apresentados a uma lista de formas truncadas, os participantes eram solicitados a identificar suas palavras-base. Para a busca realizada no Twitter, selecionamos apenas os dados em que havia 70% ou mais de concordância na escolha da palavra-base. A questão norteadora é a seguinte: será que os falantes ainda conseguem recuperar a palavra-base de truncamentos muito frequentes na língua? Por isso, esta etapa da pesquisa pretende verificar se os truncamentos em que a base foi facilmente recuperável pelos participantes dos experimentos são bastante frequentes no *Twitter*.

Foi necessário excluir um dado de truncamento que havia sido selecionado para a busca no Twitter. A forma truncada *cunha* (palavra-base *cunhado(a)*) precisou ser excluída da busca, pois os dados encontrados no Twitter eram, em sua maioria, do sobrenome “Cunha”.

As dez formas truncadas do Exercício 1 mais frequentes no Twitter foram os apresentados na Tabela 3. Calculamos a média dos resultados das três buscas realizadas para organizarmos os dados em mais e menos frequentes.

Tabela 3 – Dez truncamentos do exercício 1 mais frequentes no *Twitter*

Truncamento	<i>Twitter</i>	Concordância na escolha da palavra-base	
	Frequência (média)	Experimento <i>on-line</i>	Experimento presencial
bi	3357,7	86 %	78,3 %
aniver/niver	3090,3	100 %	100 %
refri	1238,3	100 %	95,7 %
japa	1210,0	97,7 %	95,7 %
madrugada	383,3	100 %	97,8 %
bele	354,0	95,3 %	97,8 %
maraca	300,7	88,4 %	82,6 %
oftalmo	260,0	100 %	97,8 %
profe	177,0	100 %	100 %
flagra	150,3	93 %	69,6 %

Nestes dez dados de truncamento, a concordância na escolha da palavra-base pelos participantes do experimento foi de cerca de 80%. Havia cinco dados de truncamento (*aniver/niver*, *refri*, *madrugada*, *oftalmo* e *profe*) com 100% de concordância na escolha da palavra-base (*aniversário*, *refrigerante*, *madrugada*, *oftalmologista* e *professor(a)*, respectivamente). Estes dados também estão entre os dez mais frequentes na busca no *Twitter* (em azul na tabela).

O truncamento mais frequente no *Twitter* foi *bi*, com uma frequência média de ocorrência

de 3357,7 nas três buscas. Assim como nos experimentos, a grande maioria dos dados de *bi* encontrados no *Twitter* tem como palavra-base *bissexual*. Há também dados de *bi* referindo-se a *bicampeão*. É preciso, no entanto, fazer uma ressalva sobre esse dado. Para alguns autores, *bi* não é uma forma truncada, mas sim, é o resultado de um processo morfológico em que o falante utiliza o prefixo *bi* no lugar da palavra toda. Seria, por isso, um processo diferente ao do truncamento, em que a porção da palavra que permanece após a redução não é um prefixo ou sufixo. É o caso de Martini (2010, p. 254) que exclui de sua análise formas “que se estruturam se estruturam com a manutenção integral do morfema à esquerda da forma de base”, como, por exemplo, em *odonto* < *odontologia*. Por outro lado, Belchor (2014), em sua análise com base na Morfologia Prosódica Circunscritiva, define três padrões no processo do truncamento e um deles inclui justamente estas formas que preservam o radical ou o prefixo situado mais à esquerda da palavra matriz. Neste estudo, nos aproximamos mais do que propõe Belchor (2014), entendendo o truncamento como um processo no qual parte da parte da palavra é usada no lugar do todo. Por isso, incluímos dados como *bi*, *foto* e *moto* na análise e observamos se esses dados apresentam um funcionamento diferente dos demais, de modo a discutir o *status* dessas formas no processo de truncamento.

Dados do exercício 2

No exercício 2 dos experimentos realizados em Heineck (2018), os informantes propunham formas truncadas para palavras do português. Com a busca realizada no *Twitter*, tínhamos como objetivo verificar se os truncamentos mais frequentes nesse exercício eram também os mais frequentes no *Twitter*.

Na tabela abaixo está a relação dos dez dados de truncamento do exercício 2 com maior frequência no *Twitter*.

Tabela 4 – Dez truncamentos do exercício 2 mais frequentes no *Twitter*

Truncamento	<i>Twitter</i>	Ocorrência da forma truncada	
	Frequência (média)	Experimento <i>on-line</i>	Experimento presencial
fute/futi	1296	67,4 %	87 %
juve/juvi	662,3	62,8 %	37 %
responsa	619	93 %	91,3 %
odonto	613	95,3 %	95,7 %
neura	348,7	83,7 %	50 %
bisa	244	90,7 %	91,3 %
carna	237,7	88,4 %	71,7 %
traveco	216	34,9 %	60,9 %
tranks	130,3	95,3 %	78,3 %
preju	117,7	97,7 %	93,5 %

As dez formas truncadas mais frequentes no Twitter tinham, em sua grande maioria, mais de 80% de ocorrência nos resultados dos experimentos do estudo anterior (HEINECK, 2018). Apenas *juve/juvi* e *traveco* (em rosa, na tabela) tinham uma ocorrência menor que 80% nos experimentos.

Duas formas truncadas que tinham cerca de 90% de ocorrência nos dados do estudo anterior não estão entre os dez mais frequentes no Twitter. São elas *biju* e *batera*. *Biju* foi o 12º dado mais frequente no *Twitter*, e teve uma frequência média de 109,3 que é bastante próxima da frequência de *preju*, o 10º colocado da tabela. Já em *batera*, apesar de grande parte dos participantes escolherem *batera* como a forma truncada para *bateria*, este não parece ser um truncamento muito frequente na língua, pois, na busca do Twitter, foi encontrada uma média de 11,7 dados apenas.

A alta frequência de *fute/futi* e *juve/juvi*, formas truncadas para *futebol* e *juvenil*, respectivamente, confirmaram uma tendência já verificada nos resultados dos experimentos realizados em Heineck (2018). De acordo com os dados de truncamento levantados de outros estudos acerca do fenômeno, esperávamos truncamentos trissílabos e com acréscimo de *a* para as palavras *futebol* e *juvenil* – *futiba* e *juvena*. No entanto, a grande maioria dos participantes dos experimentos preferiu manter a vogal da base e formar truncamentos dissilábicos – *fute* e *juve*. Verificamos a frequência de ocorrência de *futiba* e *juvena* no *Twitter*. Encontramos uma média de apenas 12,7 dados de *futiba* e de 50 dados de *juvena*, frequências bem menores que as encontradas para *fute/futi* (1296) e *juve/juvi* (662,3).

Outro dado interessante que merece uma maior atenção é *tosta/tusta* (média de frequência de ocorrência de 51 tokens). Nos experimentos, os informantes sugeriram *tosta* e *tusta* como formas truncadas para a palavra *tostão*. No entanto, os dados destas formas truncadas encontrados no *Twitter* não parecem ter *tostão* como palavra-base, mas sim se referem a *tostado* ou *tostada* – termo usado para torrada/misto-quente ou para se referir ao calor/sol muito quente. Em alguns exemplos encontrados, *tosta* também parece ser usado para se referir a algo ou a uma pessoa “tosca” ou “idiota”. Para compreendermos melhor os significados atribuídos a *tosta* e *tusta*, fizemos a leitura das sentenças nas quais estes dados estão inseridos. Veja alguns exemplos a seguir.

(1) “*Ainda aqui estou, após uma tosta de atum e doritos de chili :)*”

(2) “*Porque é que eu acho que vou comer uma tosta mista à 1:18 da manhã ?*”

(3) “*Que tosta. Ou fico camarão ou fico com o maior bronze da vida*”

(4) “*Tava tão bem em casa sem apanhar sol nem claridade para vir para esta tosta fds que exagero*”

(5) “*@figuei_ira @gabsoaresrr Medo de reprisar e perceber que a novela é muito tosta e estragar minha lembrança boa*”

Nos exemplos (1) e (2), *tosta* tem o sentido de *torrada* ou *misto-quente*, em (2), inclusive, *tosta* está acompanhado de *mista*, isso aparece com bastante frequência nos dados. Já nos exemplos (3) e (4), *tosta* tem o sentido de *calor* ou *sol muito forte/quente*. Por fim, no exemplo (5), *tosta* está caracterizando *novela* e parece ter o sentido de “ruim”, “de mal gosto” ou talvez “tosco”, “idiota”.

Retomando a questão norteadora desta etapa do estudo, verificamos que, de um modo geral, os dados mais frequentes no exercício 2 dos experimentos eram também frequentes no *Twitter*. Podemos supor, portanto, que os participantes do experimento optaram por essas formas truncadas talvez por já as conhecerem.

Dados do exercício 3

De modo a contribuir para a discussão acerca do status da vogal final do truncamento, decidimos verificar a frequência dos diminutivos de *foto* e *moto* nos dados do *Twitter*. *Foto* e *moto* são dados interessantes para essa discussão, pois são truncamentos femininos terminados em *o* que derivam de palavras terminadas em *a* – *fotografia* e *motocicleta*. O objetivo do exercício 3 do experimento, em que os participantes deveriam propor diminutivos para formas truncadas, era o de verificar se eles optariam por manter a vogal final do truncamento ou se resgatariam o gênero da palavra-base.

Nos resultados dos experimentos *on-line* e presencial houve uma preferência pelos diminutivos de *foto* e *moto* terminados em *o*: *fotinho* e *motinho*, como vemos na tabela 5. *Fotinho* foi o diminutivo proposto por 69,8% dos participantes do experimento *on-line* e 60,9% dos participantes do experimento presencial. *Motinho* teve uma ocorrência de 55,8% no experimento *on-line* e 43,5% no experimento presencial.

Tabela 5 – Dados de frequência dos diminutivos de *foto* e *moto* no *Twitter*

Truncamento	<i>Twitter</i>	Ocorrência da forma truncada	
	Frequência (média)	Experimento <i>on-line</i>	Experimento presencial
fotinho	406,0	69,8%	60,9%
fotinha	2306,0	23,3%	23,9%
motinho	10,0	55,8%	43,5%
motinha	144,0	30,2%	26,1%

Os resultados de frequência no *Twitter*, no entanto, não confirmaram a tendência observada nos dados dos experimentos. Os diminutivos terminados em *a* tiveram frequência maior do que os diminutivos terminados em *o*. Esses dados parecem apontar para o fato de que os falantes preferem concordar com o gênero das palavras *fotografia* e *motocicleta* ou ainda que a base destes truncamentos é transparente para os falantes.

Esses resultados de frequência confirmam a tendência mais geral observada nos resultados dos experimentos do estudo anterior. Na maior parte dos dados, observou-se a correspondência da vogal final dos diminutivos dos truncamentos com o gênero da palavra, independentemente da vogal final do truncamento.

Considerações finais

Os resultados da busca realizada no *Twitter* apontam que os truncamentos nos quais os falantes recuperavam com facilidade a base nos experimentos realizados em Heineck (2018) são também os mais frequentes entre os dados buscados no *Twitter*. Podemos supor, a partir disso, que os falantes conseguem recuperar a palavra-base mesmo de truncamentos frequentes na língua.

Na busca com os dados do exercício 2, verificamos que os dados mais frequentes dos resultados dos experimentos também são, de um modo geral, os mais frequentes na busca realizada no *Twitter*. Isso parece apontar para o fato de que a escolha dos participantes por essas formas truncadas nos experimentos não se deu de forma aleatória, mas sim por essas formas talvez já serem conhecidas pelos falantes.

Por fim, a frequência dos diminutivos de *foto* e *moto* no *Twitter* aponta para um resultado diferente ao encontrado para esses dados nos experimentos. A frequência de diminutivos terminados em *a* (*fotinha* e *motinha*), ou seja, que concordam com o gênero da palavra, foi maior do que dos diminutivos terminados em *o*. Esse resultado vai ao encontro do que foi observado nos dados dos experimentos do estudo anterior de um modo mais geral. Os participantes dos experimentos parecem, geralmente, basear a escolha do diminutivo terminado em *a* ou *o*, de acordo com o gênero da palavra, independente da vogal final da forma truncada.

Este estudo caracterizou-se como uma etapa inicial acerca da influência da frequência lexical no fenômeno do truncamento. De modo a poder confirmar os resultados aqui encontrados e formular novas hipóteses sobre o papel que a frequência exerce sobre esse fenômeno, pretende-se realizar uma busca com um maior número de dados de truncamento no *Twitter*.

REFERÊNCIAS

ARAÚJO, G. Truncamento e reduplicação no português brasileiro. *Revista de Estudos da Linguagem*. Belo Horizonte, v.10, n.1, p. 61-90, jan./jun. 2002.

BASÍLIO, M. *Teoria lexical*. 3. ed. São Paulo: Ática, 1991. (Série Princípios)

BECHARA, E. *Moderna Gramática Portuguesa*. 37. ed. rev. e ampl. Rio de Janeiro: Lucerna, 2006.

BELCHOR, A.P.V. Construções de truncamento no português do Brasil: análise estrutural à luz da Teoria da Otimalidade. Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, 2009.

BELCHOR, A.P.V. A Morfologia Prosódica Circunscritiva aplicada ao truncamento no português brasileiro. Tese (Doutorado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, 2014.

BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2 ed, 2004.

- CUNHA, C.; CINTRA, L. F. L. *Nova gramática do português contemporâneo*. Rio de Janeiro: Nova Fronteira, 1985.
- FIDELHOLTZ, J. L. Word frequency and vowel reduction in English. In: REGIONAL MEETING CHICAGO LINGUISTICS SOCIETY, 7, 1975, Chicago. Papers from... Chicago: Chicago Linguistic Society. p. 200-213, 1975.
- GONÇALVES, C. A. V. Processos morfológicos não-concatenativos: formato morfoprosódico e latitude funcional. *ALFA – Revista de Lingüística*, Araraquara, v. 48, n. 1, p. 9-27, jul./dez. 2004.
- HEINECK. Aspectos morfofonológicos do truncamento no português brasileiro. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul, 2018.
- HUBACK, A. P. A interferência da frequência em fenômenos linguísticos. *Revista Delta*, 29:1, p. 79-94, 2013.
- LIMA, B. C. Realização fonética de acrônimos no português brasileiro: uma abordagem morfofonológica através da Teoria da Otimalidade. Tese (Doutorado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, 2014.
- MARTINI, L. D. *Morfologia Prosódica no Português Brasileiro*. Belo Horizonte: Faculdade de Letras/UFMG, 2010.
- ROCHA, L. C. *Estruturas morfológicas do português*. Belo Horizonte: Ed. UFMG, 1998.
- SANDMANN, A. *Morfologia lexical*. São Paulo: Contexto, 1992. (Coleção Repensando a Língua Portuguesa)
- SCHER, A. P. A study of truncated nominal forms in Brazilian Portuguese: their derivation and their relation to nonverbal form classes. In: Mary A. Kato; Francisco Ordóñez.(Org.). *The Morphosyntax of Portuguese and Spanish in Latin America*. 1ª ed. New York: Oxford University Press, 2016, p. 332-353.
- SCHER, A. P. Formas truncadas em português brasileiro e espanhol peninsular: descrição preliminar. *ReVEL*, edição especial n. 5, 2011.
- SILVA, T. C. *Dicionário de fonética e fonologia*. São Paulo: Contexto, 2011.
- VILELA, A. C.; GODOY, L. & SILVA, T. C. Truncamento no PB - para uma melhor compreensão do fenômeno. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 14, n. 1, p. 149-179, 2006.
- VILLALVA, A. *Morfologia do Português*. Lisboa: Universidade Aberta, 2007.