

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**TRUNCAMENTO EM *CORPUS* DO PORTUGUÊS BRASILEIRO: DESCRIÇÃO,
ANÁLISE E INVESTIGAÇÃO EXPERIMENTAL**

DÉBORA HEINECK

PORTO ALEGRE

2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ESTUDOS DA LINGUAGEM
FONOLOGIA E MORFOLOGIA

**TRUNCAMENTO EM *CORPUS* DO PORTUGUÊS BRASILEIRO: DESCRIÇÃO,
ANÁLISE E INVESTIGAÇÃO EXPERIMENTAL**

DÉBORA HEINECK

Tese de Doutorado em Estudos da Linguagem, apresentada como requisito parcial para a obtenção do título de Doutora pelo Programa de Pós-graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientação: Prof. Dr. Luiz Carlos Schwandt

PORTO ALEGRE

2023

CIP - Catalogação na Publicação

Heineck, Débora
TRUNCAMENTO EM CORPUS DO PORTUGUÊS BRASILEIRO:
DESCRIÇÃO, ANÁLISE E INVESTIGAÇÃO EXPERIMENTAL /
Débora Heineck. -- 2023.
144 f.
Orientador: Luiz Carlos Schwindt.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2023.

1. Truncamento. 2. Morfofonologia. 3. Frequência
lexical. 4. Processamento morfológico. I. Schwindt,
Luiz Carlos, orient. II. Título.

Débora Heineck

TRUNCAMENTO EM *CORPUS* DO PORTUGUÊS BRASILEIRO: DESCRIÇÃO,
ANÁLISE E INVESTIGAÇÃO EXPERIMENTAL

Tese de Doutorado em Estudos da Linguagem, apresentada como requisito parcial para a obtenção do título de Doutora pelo Programa de Pós-graduação em Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 23 de junho de 2023.

Resultado: Aprovada

Luiz Carlos da Silva Schwindt – Orientador
Universidade Federal do Rio Grande do Sul

Rafael Dias Minussi
Universidade Federal de São Paulo

Camila Witt Ulrich
Universidade Federal do Pampa – Campus Jaguarão

Felipe Bilharva da Silva
Universidade Federal do Rio Grande do Sul

AGRADECIMENTOS

Primeiramente, quero agradecer à minha família. Meus pais foram sempre meu porto seguro, me dão alento nos momentos difíceis e vibram com as minhas conquistas. O Thiago que me ajuda na programação em R, na coleta de dados, nos problemas tecnológicos e que jamais deixa de acreditar em mim. Obrigada por todo amor, carinho e compreensão! A Irene sempre soube respeitar e acolher minhas angústias, como uma segunda mãe! E a Renata, que foi e sempre será a minha irmã do coração, o meu lar longe de casa. Amo vocês!

Agradeço ao meu orientador, prof. Luiz Carlos Schwindt, que no segundo semestre da graduação, viu meu olho brilhar pela fonologia e morfologia e me convidou a participar do Círculo Linguístico: Fonologia e Morfologia, atual MorPhon. Sou muito grata por ter tido a oportunidade de ser aluna e orientanda de uma pessoa tão brilhante e inspiradora. O olhar preciso e cuidadoso do prof. Luiz ao meu trabalho e à minha caminhada fizeram toda a diferença!

Agradeço à UFRGS pela oportunidade de um ensino gratuito e de qualidade na graduação, mestrado e doutorado. Estendo o agradecimento ao PPG-Letras da UFRGS pela oportunidade e pelo ambiente de aprendizado. Tenho muito orgulho em fazer parte desta instituição.

Aos membros da banca, professores Camila Ulrich, Felipe Bilharva e Rafael Minussi, agradeço imensamente pela disponibilidade em ler e contribuir para o crescimento deste trabalho. Ao professor Leandro Lara, que acompanhou o andamento da tese, sou muito grata por todas as contribuições e apontamentos.

Dedico um agradecimento especial à professora Gisela Collischonn, que me introduziu ao mundo da pesquisa científica, me apresentou às belezas da fonologia e tão humildemente compartilhou comigo seu vasto conhecimento sobre linguística. O início da minha caminhada foi essencial para que eu chegasse até aqui.

A todos os colegas do MorPhon por todas as trocas, discussões, contribuições e por todo o aprendizado. O grupo foi muito importante para o meu crescimento como pesquisadora e, a partir dele, fiz amizades que levo com carinho no coração!

Agradeço a todos os participantes dos experimentos, por seu tempo e paciência em responder aos formulários *online*. Vocês foram fundamentais para esse trabalho.

Aos colegas, alunos, coordenação e direção do Colégio Santa Inês, que acompanharam minha jornada e sempre me apoiaram para que essa conquista fosse possível.

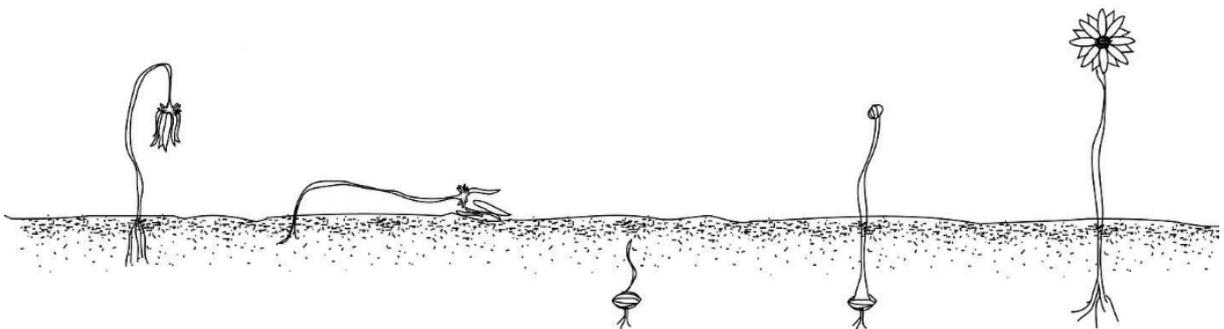
Ao Samuel Oliveira, minha alma gêmea em forma de amigo, por estar ao meu lado sempre, por me acompanhar, me escutar e me compreender! Dear, foi muito importante para mim te ver crescer e te tornar esse pesquisador e profissional incrível! Tu é minha inspiração!

A todos meus amigos de Porto Alegre e Arroio do Meio, por compreenderem minhas ausências, por torcerem pelo meu sucesso e por tornarem mais leves até os momentos mais pesados vida.

EPÍGRAFE

*this is the recipe of life
said my mother
as she held me in her arms as i wept
think of those flowers you plant
in the garden each year
they will teach you
that people too must wilt
fall
root
rise
in order to bloom*

- rupi kaur



RESUMO

Este trabalho tem por objetivo descrever o funcionamento do processo de truncamento no português brasileiro, delimitar suas regularidades e, conseqüentemente, dar continuidade a estudos a respeito do processo. Pretendemos contribuir para estudos sobre o papel de fatores fonológicos e morfológicos e da frequência lexical para o processo, bem como para a discussão sobre como se dá o processamento morfológico de processos como o truncamento.

Para tanto, foi realizado um levantamento de dados a partir de estudos anteriores sobre o fenômeno (ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014; SCHER, 2018), dos resultados obtidos em Heineck (2018) e do registro de dados de redes sociais, programas de televisão, entre outros. Esses dados foram organizados e decodificados a partir de critérios fonológicos e morfológicos, como acento e o número de sílabas, foi registrada a frequência no Corpus do Português (subcorpus *Web/Dialects*) da forma truncada e de sua palavra-base e foram acrescentados exemplos de uso de cada dado. O nosso *corpus* de dados de truncamento conta com 239 itens no total.

Os dados de Heineck (2018) foram submetidos a uma análise descritiva e inferencial com o objetivo de observar o papel do número de sílabas para o processo do truncamento. Os resultados apontaram para uma tendência de as formas truncadas serem maiores quando há um aumento no número de sílabas da palavra-base. Esse resultado confirma os resultados de Heineck (2018) e corrobora o padrão encontrado por Martini (2010) na formação de truncamentos: palavras dissílabas geralmente formam truncamentos também dissílabos, palavras trissílabas geralmente formam truncamentos dissílabos e palavras polissílabas tendem a formar truncamentos trissílabos.

De modo a investigar o papel da frequência lexical no processo de truncamento, o número de ocorrência de dados dos experimentos realizados por Heineck (2018) foi verificada na rede social *Twitter*. 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.

Todos os 239 dados que compõem o nosso *corpus* foram verificados quanto a sua frequência de ocorrência no Corpus do Português (subcorpus *Web/Dialects*) e o resultado desta busca foi comparada com os resultados do estudo com dados do *Twitter*. Verificamos que os truncamentos mais frequentes do nosso *corpus* tinham como base palavras frequentes, mas algumas das palavras-base mais frequentes desse *corpus* têm formas truncadas com poucas ou até zero ocorrências. Pudemos inferir, nesse sentido, que truncamentos frequentes parecem ter origem em palavras também frequentes, mas que nem toda palavra muito frequente irá gerar truncamentos frequentes. Na comparação com os resultados de Heineck (2021) com os resultados da busca no *Web/Dialects*, verificamos que alguns dos dados mais frequentes no *Twitter* não eram frequentes no *Web/Dialects*. Isso pode ocorrer, pois o *Twitter* é um ambiente mais informal e mais propício para o uso de formas como o truncamento do que os sites que compõe o banco de dados do *Web/Dialects*.

Para contribuir com a discussão sobre processamento morfológico do truncamento, desenvolvemos uma tarefa de decisão lexical com *priming*, em que avaliamos o índice de acertos e o tempo de reação nas respostas de 62 falantes de português brasileiros para os estímulos apresentados. Avaliamos a influência de variáveis como acento, número de sílabas e a relação entre *prime* e alvo. O experimento era composto de 60 truncamentos e 30 não palavras, que serviam como itens distratores. O experimento foi dividido em três listas e aplicado na plataforma Psytoolkit 3.3.4. As variáveis predictoras analisadas — relação entre *prime* e alvo, características morfofonológicas da base e do truncamento — não apresentaram diferenças significativas quanto ao índice de acertos. Em relação ao tempo de reação, a análise confirmou que o tempo de reação é menor quando há relação entre palavra-base e o truncamento. Isso parece confirmar que o processamento da forma truncada é mais rápido quando o falante o relaciona com sua palavra-base. As características morfofonológicas tanto da palavra-base quanto do truncamento não apresentaram diferenças significativas em relação ao tempo de reação. É importante salientar, no entanto, que os dados apresentaram distribuição não normal e, portanto, não foi possível aplicar testes paramétricos em nossa análise. Nessa perspectiva, não se descarta encontrar condicionamento das variáveis morfofonológicas, para as quais não se identificou papel significativo neste estudo.

Palavras-chave: truncamento; morfofonologia; frequência lexical; processamento morfológico.

ABSTRACT

This work aims to describe the operation of the truncation process in Brazilian Portuguese, delimit its regularities and, consequently, continue studies on the process. We intend to contribute to studies on the role of phonological and morphological factors and lexical frequency for the process, as well as to the discussion about how the morphological processing of processes such as truncation takes place.

To do that, a data collection was carried out from previous studies on the phenomenon (ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014; SCHER, 2018), from the results obtained in Heineck (2018) and from the registration of data from social networks, television programs, among others. These data were organized and decoded based on phonological and morphological criteria, such as word stress and the number of syllables. The frequency in the Corpus do Português (subcorpus Web/Dialects) of the truncated form and its base word and examples of use of each data were registered. Our corpus of truncation data has 239 items in total.

Heineck's (2018) data were subjected to a descriptive and inferential analysis in order to observe the role of the number of syllables in the truncation process. The results pointed to a tendency for the truncated forms to be larger when there is an increase in the number of syllables in the base word. This result confirms the results of Heineck (2018) and corroborates the pattern found by Martini (2010) in the formation of truncations: disyllable words generally form disyllable truncations, trisyllable words generally form disyllable truncations, and polysyllable words tend to form trisyllable truncations.

In order to investigate the role of lexical frequency in the truncation process, the number of data occurrence of the experiments carried out by Heineck (2018) was verified on the social network Twitter. The results of the search performed on Twitter seem to confirm some of our hypotheses: (a) the speakers can recover the base word even from frequent truncations in the language and (b) the most frequent data in the experiments carried out in the Master's Thesis are also the most frequent in the search on Twitter, which raises the assumption that the speakers already knew these forms. Finally, the frequency of the diminutives of *foto* and *moto* points to a different result to that found in the experiments but follows a more general trend found for other data in the experiments. When forming the diminutive of a truncated form, 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.

To contribute to the discussion on the morphological processing of truncation, we developed a lexical decision task with priming, in which we evaluated the rate of correct answers and reaction time in the responses of 62 Brazilian Portuguese speakers to the stimuli presented. We evaluated the influence of variables such as word stress, number of syllables and the relationship between prime and target. The experiment was composed of 60 truncated words and 30 non-words, which served as distracting items. The experiment was divided into three lists and applied on the Psytoolkit 3.3.4 platform. The predictive variables analyzed — relationship between prime and target, morphophonological characteristics of the base and of the truncated word — did not present significant differences regarding the rate of correct answers. Regarding the reaction time, the analysis confirmed that the reaction time is shorter when there is a relationship between the base word and the truncation. This seems to confirm that processing the truncated form is faster when the speaker relates it to their base word. The morphophonological characteristics of both the base word and the truncation did not present significant differences in relation to the reaction time. It is important to note, however, that the data presented non-normal distribution and, therefore, it was not possible to apply parametric tests in our analysis. In this perspective, it is not ruled out to find conditioning of morphophonological variables, for which no significant role was identified in this study.

Kew-words: truncation; morfophonology; lexical frequency; morphological processing.

LISTA DE TABELAS

Tabela 1 – Média, desvio padrão e intervalos de confiança das variáveis <i>monossílabo</i> , <i>dissílabo</i> e <i>trissílabo</i> do exercício 1.....	50
Tabela 2 – Média, desvio padrão e intervalos de confiança das variáveis <i>dissílabo</i> , <i>trissílabo</i> e <i>polissílabo</i> do exercício 2.....	53
Tabela 3 – Proporção do número de sílabas da palavra-base em relação ao número de sílabas do truncamento	57
Tabela 4 – Proporção do acento da palavra-base em relação ao acento do truncamento	59
Tabela 5 – Dados do exercício 1 com variação de frequência de ocorrência maior que 20% em relação à média da frequência nas três buscas realizadas no <i>Twitter</i>	71
Tabela 6 – Dados do exercício 2 com variação de frequência de ocorrência maior que 30% em relação à média da frequência nas três buscas realizadas no <i>Twitter</i>	72
Tabela 7 – Dez truncamentos do exercício 1 mais frequentes no <i>Twitter</i>	73
Tabela 8 – Dez truncamentos do exercício 2 mais frequentes no <i>Twitter</i>	75
Tabela 9 – Dados de frequência dos diminutivos de <i>foto</i> e <i>moto</i> no <i>Twitter</i>	77
Tabela 10: Distribuição da frequência em <i>tokens</i> dos dados de truncamento	80
Tabela 11: Dados de truncamento com mais de 10.000 ocorrências.....	80
Tabela 12: Distribuição da frequência em <i>tokens</i> dos dados de palavras-base	81
Tabela 13: Dados de palavras-base com mais de 100.000 ocorrências.....	81
Tabela 14 – Dez truncamentos do exercício 1 mais frequentes no <i>Twitter</i> e suas frequências no <i>Web/Dialects</i>	82
Tabela 15 – Dez truncamentos do exercício 2 mais frequentes no <i>Twitter</i> e suas frequências no <i>Web/Dialects</i>	83
Tabela 16 – Condições para a seleção de dados para a tarefa de decisão lexical com <i>priming</i>	90
Tabela 17 – Perfil dos respondentes da tarefa de decisão lexical com <i>priming</i> por lista	91
Tabela 18 – Distribuição do índice de acerto em relação à variável <i>relação entre prime e alvo</i>	95
Tabela 19 – Distribuição do índice de acerto em relação à variável <i>características morfofonológicas da palavra-base</i>	95

Tabela 20 – Distribuição do índice de acerto em relação à variável <i>características morfofonológicas do truncamento</i>	96
Tabela 21 – Medianas do tempo de reação em relação à variável <i>características morfofonológicas da palavra-base</i>	99
Tabela 22 – Medianas do tempo de reação em relação à variável <i>características morfofonológicas do truncamento</i>	101

LISTA DE GRÁFICOS

Gráfico 1 – Distribuição dos dados das variáveis <i>monossílabo</i> , <i>dissílabo</i> e <i>trissílabo</i> do exercício 1.....	49
Gráfico 2 – Intervalo de confiança das variáveis <i>monossílabo</i> , <i>dissílabo</i> e <i>trissílabo</i> do exercício 1	51
Gráfico 3 – Distribuição dos dados das variáveis <i>dissílabo</i> , <i>trissílabo</i> e <i>polissílabo</i> do exercício 2	52
Gráfico 4 – Frequência de ocorrência dos dados do exercício 1 nas três buscas realizadas no <i>Twitter</i>	70
Gráfico 5 – Frequência de ocorrência dos dados do exercício 2 nas três buscas realizadas no <i>Twitter</i>	71
Gráfico 6 – Frequência de ocorrência dos dados do exercício 3 nas três buscas realizadas no <i>Twitter</i>	72
Gráfico 7 – Índice de acerto dos dados da tarefa de decisão lexical com <i>priming</i>	94
Gráfico 8 – Distribuição dos dados da variável relação entre <i>prime</i> e alvo em relação ao tempo de reação	98
Gráfico 9 - Distribuição dos dados da variável <i>características morfofonológicas da palavra-base</i> em relação ao tempo de reação	100
Gráfico 10 - Distribuição dos dados da variável <i>características morfofonológicas do truncamento</i> em relação ao tempo de reação.....	101

LISTA DE FIGURAS

Figura 1 - Esquema de hierarquia de restrições de fidelidade e marcação no truncamento.....	33
Figura 2 - Representação da inserção de material fonológico à base do truncamento.....	35
Figura 3 - Diagrama do processo do truncamento proposto por Scher (2011).....	40
Figura 4 – Exemplo do exercício 1 do experimento presencial	44
Figura 5 – Exemplo do exercício 2 do experimento presencial	45
Figura 6 – Exemplo do exercício 3 do experimento presencial	46
Figura 7 – Distribuição dos informantes por estados brasileiros na tarefa de decisão lexical com <i>priming</i>	92

SUMÁRIO

1 INTRODUÇÃO	18
PARTE 1.....	21
O PROCESSO DO TRUNCAMENTO NO PORTUGUÊS BRASILEIRO	21
2 ESTUDOS SOBRE O TRUNCAMENTO NO PORTUGUÊS.....	22
2.1 ANÁLISES DE GRAMÁTICAS TRADICIONAIS	22
2.2 ANÁLISES DESCRITIVAS	24
2.2.1 Morfologia Prosódica (MP)	24
2.2.1.1 Araújo (2002)	26
2.2.1.2 Gonçalves (2004).....	30
2.2.1.3 Gonçalves e Vasquez (2004)	31
2.2.1.4 Martini (2010)	33
2.2.1.5 Belchor (2014).....	35
2.2.2 Morfologia Distribuída (DM).....	37
2.2.2.1 Scher (2011)	38
2.2.2.2 Scher (2016)	41
3 CONSTITUIÇÃO DO <i>CORPUS</i>	43
3.1 HEINECK (2018)	43
3.1.1 Levantamento de dados	43
3.1.2 Principais resultados de Heineck (2018)	46
3.2 PAPEL DO NÚMERO DE SÍLABAS	47
3.2.1 O número de sílabas no truncamento.....	47
3.2.2 Metodologia.....	48
3.2.3 Resultados	49
3.2.3.1 Exercício 1	49
3.2.3.2 Exercício 2	52

3.2.4 Conclusões gerais do estudo sobre o papel do número de sílabas no truncamento	54
3.2 REORGANIZAÇÃO DO <i>CORPUS</i>	55
3.4.1 Número de sílabas	57
3.4.2 Acento	59
3.4.3 Frequência	59
3.4.4 Palavra-base formada por duas palavras morfológicas e/ou duas palavras fonológicas	60
3.4.5 Acréscimo de -a	62
3.4.6 Acréscimo de -s	63
3.4.7 Origem do dado e exemplo de uso	64
PARTE 2	65
PAPEL DA FREQUÊNCIA LEXICAL NO TRUNCAMENTO	65
5 FREQUÊNCIA LEXICAL	66
5.1 FONOLOGIA DE USO	66
6 ESTUDO DE FREQUÊNCIA COM DADOS DO <i>TWITTER</i> E DO <i>CORPUS</i> DO PORTUGUÊS	68
6.1 HEINECK (2021)	68
6.1.1 Seleção dos dados	68
6.1.2 Busca no <i>Twitter</i>	69
6.1.3 Principais resultados	70
6.1.3.1 Dados do exercício 1	73
6.1.3.2 Dados do exercício 2	74
6.1.3.3 Dados do exercício 3	77
6.1.4 Conclusões gerais do estudo sobre frequência no <i>Twitter</i>	78
6.2 FREQUÊNCIA COM DADOS DO <i>CORPUS</i> DO PORTUGUÊS	78
6.2.1 Principais resultados	79
6.2.2 Comparação com dados do <i>Twitter</i>	82

6.2.3 Conclusões gerais do estudo sobre frequência no Corpus do Português	84
PARTE 3.....	86
INVESTIGAÇÃO EXPERIMENTAL	86
7 TAREFA DE DECISÃO LEXICAL COM <i>PRIMING</i>	87
7.1 SELEÇÃO DOS DADOS.....	87
7.2 CONDIÇÕES DO EXPERIMENTO.....	88
7.3 PARTICIPANTES	90
7.4 ANÁLISE DOS DADOS.....	92
7.4.1 Índice de acerto.....	93
7.4.1.1 Relação entre <i>prime</i> e alvo	94
7.4.1.2 Características morfofonológicas da palavra-base	95
7.4.1.3 Características morfofonológicas do truncamento	96
7.4.2 Tempo de reação.....	96
7.4.2.1 Relação entre <i>prime</i> e alvo	97
7.4.2.2 Características morfofonológicas da palavra-base	99
7.4.2.3 Características morfofonológicas do truncamento	100
7.5 Conclusões gerais da tarefa de decisão lexical com <i>priming</i>	102
8 CONSIDERAÇÕES FINAIS.....	103
9 REFERÊNCIAS BIBLIOGRÁFICAS	108
10 ANEXOS	112
10.1 APROVAÇÃO DA PESQUISA PELO COMITÊ DE ÉTICA	112
10.2 CONVITE PARA PARTICIPAÇÃO EM PESQUISA	122
10.3 TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DO EXPERIMENTO DE PROCESSAMENTO MORFOLÓGICO	122
10.4 DADOS UTILIZADOS NO EXPERIMENTO.....	124
10.5 MODELO DE <i>SCRIPT</i> UTILIZADO NA TAREFA DE DECISÃO LEXICAL COM <i>PRIMING</i>	128

10.6 ITENS COM ZERO OCORRÊNCIAS NO <i>WEB/DIALECTS</i> EM NOSSO <i>CORPUS</i> DE DADOS DE TRUNCAMENTOS	130
10.7 LISTA DE DADOS QUE COMPÕEM NOSSO <i>CORPUS</i> DE DADOS DE TRUNCAMENTO.....	131
10.8 EXPERIMENTO ONLINE (HEINECK, 2018)	136
10.9 EXPERIMENTO PRESENCIAL (HEINECK, 2018).....	139
10.10 LISTA DE PALAVRAS DE BUSCA REALIZADA NO TWITTER (HEINECK, 2021).....	143

1 INTRODUÇÃO

O truncamento é um processo de formação de palavras que se diferencia de outros processos tradicionais da língua portuguesa, pois, ao invés de acrescentar conteúdo morfológico à base (como nos processos de prefixação e sufixação), parte da palavra é subtraída. Nos exemplos em (1) vemos que essa supressão pode ocorrer no início da palavra (como em 1a) ou no final da palavra (como em 1c).

- (1) a. paranoia > noia
- b. professor > sor
- c. publicidade > publi
- d. burocracia > burocra

Devido a essas características que diferenciam o truncamento dos processos de formação de palavras tradicionais do português, o fenômeno foi visto como idiossincrático e assistemático por alguns gramáticos e pesquisadores (CUNHA&CINTRA, 1985; BECHARA, 2006; BASILIO, 1991; SANDMANN, 1992). No entanto, estudos mais recentes procuram provar que o truncamento é um fenômeno sistemático. Gonçalves (2004) define o truncamento como um processo não concatenativo de formação de palavras. Dentro dessa categoria estão outros processos que não se encaixam no que seria o padrão da língua portuguesa, como, por exemplo, a reduplicação, a hipocorização, a mesclagem lexical e a siglagem. As análises sistemáticas do truncamento são feitas com base em diferentes teorias: Teoria da Otimidade (ARAÚJO, 2002; GONÇALVES&VASQUEZ, 2004), Morfologia Prosódica (GONÇALVES, 2004; MARTINI, 2010; BELCHOR, 2014) e Morfologia Distribuída (SCHER, 2001; SCHER, 2016).

Esta pesquisa tem como objetivo descrever o seu funcionamento, delimitar suas regularidades e comparar com estudos anteriores sobre o processo no português brasileiro. Este estudo visa, portanto, a realizar uma análise sistemática e que busca regularidades no processo estudado e, por esse motivo, pode contribuir para a caracterização, já feita em trabalhos anteriores (ARAÚJO, 2002; GONÇALVES, 2004A, 2004B; MARTINI, 2010; SCHER, 2011, 2016; entre outros), do truncamento como um processo regular e sistemático de formação de palavras.

Este trabalho também visa a observar o papel da frequência lexical no fenômeno do truncamento, ou seja, verificar se a frequência de uso da palavra-base e/ou do truncamento exerce influência no processo. Não encontramos, na literatura, trabalhos sobre processos não

concatenativos que levassem em consideração a frequência lexical em suas análises. Este trabalho, portanto, pode contribuir para uma discussão dos efeitos de frequência nesse tipo de processo de formação de palavras, bem como contribuir para a discussão já bastante forte sobre o papel da frequência lexical em fenômenos linguísticos.

Pretende-se, ainda, contribuir para estudos acerca do processamento morfológico de fenômenos como o truncamento por falantes do português brasileiro. A partir de uma tarefa de decisão lexical com *priming*, objetivamos observar de que modo os falantes processam as palavras truncadas e se esse processamento é facilitado por alguns fatores como, por exemplo, o reconhecimento da palavra-base e a frequência da palavra-base e do truncamento.

Esta pesquisa, em síntese, pode contribuir para os estudos acerca de processos não concatenativos do português, em especial, o truncamento, ampliando a discussão sobre o funcionamento do fenômeno, bem como auxiliar na compreensão da função das formas truncadas no âmbito social. Este trabalho, em especial, irá contribuir para estudos sobre o papel da frequência lexical e sobre processamento morfológico de processos como o truncamento. É importante salientar que pretendemos oferecer uma análise descritiva e empírica do truncamento e, nesse estudo, não temos como objetivo construir uma proposta de modelagem do fenômeno em teorias formais.

Como objetivos específicos desta pesquisa, podemos mencionar:

- compor um *corpus* de dados atuais de truncamentos que possa contribuir para os *corpora* de dados deste processo já existentes;
- investigar o papel da tonicidade e do tamanho da palavra-base na formação de truncamentos;
- discutir o papel da frequência lexical no processo do truncamento, de modo a verificar se a palavra, mais do que sequências sonoras ou as estruturas morfológicas, é que está sujeita ao processo de truncamento;
- investigar de que modo se dá o processamento morfológico do truncamento por falantes do português brasileiro.

De modo a alcançar estes objetivos, tomamos como base para a pesquisa as seguintes questões norteadoras:

- a) de que forma fatores como o número de sílabas e acento influenciam na formação de truncamentos?
- b) o truncamento é um processo produtivo de formação de palavras?

- c) a frequência da palavra-base tem influência na produção de formas truncadas? A frequência da forma truncada depende da frequência de sua palavra-base ou é possível que um truncamento seja mais frequente que sua palavra de origem?
- d) de que forma os falantes processam as formas truncadas? o processamento é mais rápido e eficiente quando os falantes relacionam o truncamento com sua palavra-base?

Este trabalho está organizado em três partes.

A parte 1 é composta pelos capítulos 2, 3 e 4. O capítulo 2 constitui-se de uma revisão bibliográfica dos estudos acerca do truncamento no português brasileiro. O capítulo 3 trata da composição do *corpus* de dados de truncamento que teve como ponto de partida o trabalho realizado em Heineck (2018). O capítulo 4 trata de um estudo acerca do papel do número de sílabas tanto da palavra-base quanto da forma truncada para a formação de truncamentos.

A parte 2 trata do papel da frequência lexical no truncamento e é composta dos capítulos 5 e 6. No capítulo 5, tratamos da Fonologia de Uso proposta por Bybee (2004) e do papel da frequência lexical em processos linguísticos. O capítulo 6 trata de um estudo de frequência com dados de truncamento realizada por Heineck (2021).

Na parte 3, descrevemos a metodologia utilizada na construção de uma tarefa de decisão lexical com *priming* e analisamos seus resultados de forma a contribuir para as discussões acerca de processamento morfológico de processos como o truncamento.

Por fim, o capítulo 8 apresenta nossas considerações finais e os próximos passos deste trabalho. Na sequência, estão nossas referências bibliográficas e os anexos deste trabalho.

PARTE 1

O PROCESSO DO TRUNCAMENTO NO PORTUGUÊS BRASILEIRO

2 ESTUDOS SOBRE O TRUNCAMENTO NO PORTUGUÊS

2.1 ANÁLISES DE GRAMÁTICAS TRADICIONAIS

Tradicionalmente, o truncamento é entendido, em gramáticas e em alguns manuais de morfologia, como um fenômeno idiossincrático do português e, por isso, geralmente, não se encontra uma descrição do seu funcionamento, mas apenas listas de exemplos. Nas gramáticas tradicionais, o truncamento é geralmente equiparado a processos como o de abreviação, como veremos na sequência.

Cunha e Cintra (1985) tratam de compostos greco-latinos que sofrem um processo de encurtamento, como em *auto* < *automóvel*, *foto* < *fotografia*, *moto* < *motocicleta*, *pneu* < *pneumático*, *quilo* < *quilograma* etc. De acordo com os autores, o processo, que é denominado como *Abreviação vocabular*, não prejudica a compreensão, pois a forma abreviada assume o sentido da forma plena. Análises linguísticas mais recentes (GONÇALVES, 2004; MARTINI, 2010) vão de encontro ao que afirma Cunha e Cintra, pois destacam que as formas truncadas apresentam diferenças pragmáticas em relação a sua forma de base.

Uma possível justificativa para a ocorrência do processo de truncamento, de acordo com os gramáticos, é a economia linguística, pois “economizar tempo e palavras” seria “uma tendência geral do mundo de hoje”, consequência do ritmo acelerado da vida intensa de nossos dias, que nos obriga a uma “elocução mais rápida” (CUNHA E CINTRA, 1985, p.114).

Bechara (2006) também denomina o truncamento de abreviação e o define, de forma muito sucinta, como o “emprego de uma parte da palavra pelo todo” (p.309). O autor cita poucos exemplos como *extra* para *extraordinário* ou *extrafino*, e *foto* para *fotografia*. Afirma ainda que o processo é comum tanto na fala coloquial, quanto na linguagem cuidada, por “breviedade de expressão”. O autor define como um caso especial da abreviação o processo de criação de palavra através da leitura das letras que compõem siglas, processo que Gonçalves (2004) denomina *siglagem*.

As definições de Cunha & Cintra e Bechara se assemelham ao denominar o truncamento como *abreviação* e por usarem de uma ideia de economia de tempo e de palavras como justificativa para sua ocorrência.

Nos manuais de linguística, os autores diferenciam de forma mais acurada os processos de encurtamento, separando o truncamento de processos como acronímia e abreviação, por exemplo. No entanto, esses estudos ainda descrevem os processos como imprevisíveis, sem realizar a descrição do seu funcionamento.

Basilio (1991) cita alguns exemplos de formas truncadas como *portuga* < *português*, *boteco* < *botequim*, *delega* < *delegado*, etc., e afirma que eles são resultado de um processo de encurtamento, a que chama de *redução* ou *abreviação*. A autora não identifica um padrão nesse processo e afirma que, na maioria dos casos, a parte a ser suprimida é imprevisível. Basilio trata ainda de truncamentos advindos de bases compostas, como em *mini* < *minissaia*, *micro* < *microcomputador* e *vídeo* < *videocassete*. Nesses casos, a redução se dá pelos constituintes da composição. Por isso, a autora vê regularidade na redução desses exemplos. Em relação à semântica, a autora vai ao encontro do que outros linguistas verificam em estudos posteriores e propõe que a palavra truncada é sinônima da sua matriz e é usada, geralmente, em um estilo mais coloquial.

De forma semelhante a Basilio, Sandmann (1992) distingue dois tipos de abreviação: palavras do tipo *cerva* e palavras do tipo *maxi*. No grupo de palavras do tipo *maxi*, o autor identifica regularidade no processo de encurtamento, pois a estrutura da palavra complexa é levada em consideração, como ocorre em *maxi* que pode ser abreviação para *maxissaia* ou *maxidesvalorização*. Já no primeiro grupo, formado por palavras do tipo *cerva*, o autor não vê a mesma regularidade, pois no caso de *cerveja*, por exemplo, não há, sincronicamente, mais de um morfema e, por isso, a abreviação *cerva* é considerada arbitrária, visto que o corte não se dá no limite entre constituintes de uma palavra morfologicamente complexa.

Villalva (2007) trata do truncamento em um capítulo intitulado “Processos não morfológicos na Formação de Palavras”. Neste capítulo, a autora afirma que tratará de fenômenos que não são do domínio da morfologia, mas produto da criatividade dos falantes. Diferentemente de alguns autores, Villalva faz uma diferenciação dos processos de truncamento, acronímia e formação de siglas e abreviaturas.

Em relação ao processo de truncamento a autora afirma que

(...) é um processo de redução de uma palavra, que elimina uma sequência, geralmente no final da palavra, e pode associar à sequência truncada um índice temático geralmente distinto do índice temático da base, sendo imprevisivelmente *-a* ou *-o*.

(VILLALVA, 2007, p.50)

A autora afirma que, de um modo geral, as formas truncadas são masculinas e não admitem contraste de gênero. Quando admitem esse contraste, ele é feito sintaticamente (*um comuna* / *uma comuna*). Ou, ainda, o truncamento preserva o valor de gênero da forma de base (*um heli(cóptero)* / *uma expo(sição)*). De forma diferente da afirmação de Villalva, os

resultados a que chegamos em Heineck (2018) sinalizam para dados de truncamentos que são sempre femininos, como é o caso de *sapata* e *satisfa*.

Assim como Basilio (1991) e Sandmann (1992), Villalva percebe regularidade em formas como *foto* < *fotografia*, *metro* < *metropolitano*, pois, nesses casos, “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.” (p. 51). A autora apresenta, ainda, alguns exemplos de hipocorísticos que, de acordo com sua análise, são formados por truncamento, como *Quim* < *Joaquim*, *Nando* < *Fernando*, por exemplo.

Apesar de os autores dos manuais de morfologia admitirem que há regularidade em formas truncadas derivadas de compostos, como *foto* < *fotografia* e *mini* < *minissaia*, não há grande inovação em relação às gramáticas tradicionais na descrição do truncamento. Os autores ainda veem processos como o truncamento, a siglagem e a hipocorização como marginais, e não realizam uma análise formal do funcionamento destes fenômenos.

2.2 ANÁLISES DESCRITIVAS

Esta seção dedica-se à apresentação de propostas de autores que realizaram análises do truncamento com o intuito de sistematizar o processo e refutar a ideia de que sua formação é irregular e imprevisível. Os estudos aqui apresentados têm como base duas teorias morfológicas: a Morfologia Prosódica (MP) e a Morfologia Distribuída (DM), que serão brevemente descritas a seguir.

2.2.1 Morfologia Prosódica (MP)

A Morfologia Prosódica é um modelo não linear, proposto por McCarthy & Prince (1986), para tratar de fenômenos que se encontram na interface fonologia-morfologia. São três os princípios básicos da Morfologia Prosódica:

- (i) Hipótese Básica da Morfologia Prosódica:
Templates são definidos em termos de unidades prosódicas, tais como mora (μ), sílaba (δ), pé (Σ) e palavra fonológica (ω);
- (ii) Condições de Satisfação do Molde:
Fenômenos morfológicos satisfazem um molde específico, que pode ser determinado por princípios universais da Prosódia ou por princípios de formação de cada língua específica;

(iii) Circunscrição Prosódica:

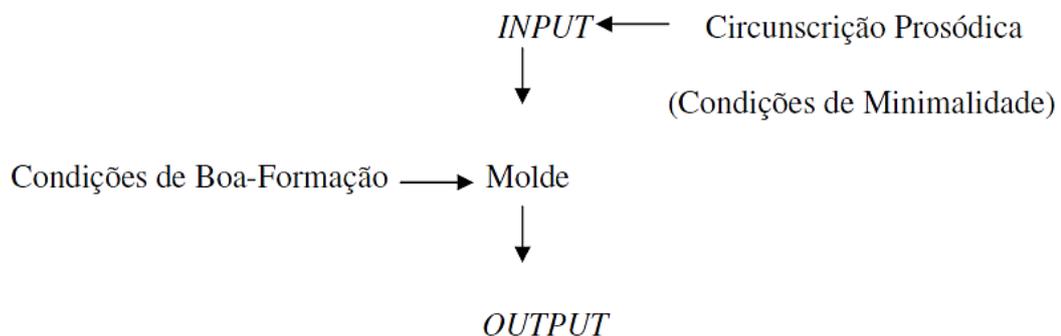
O domínio sobre o qual certas operações morfológicas atuam pode ser mapeado por primitivos prosódicos.

Fonte: Martini (2010)

No primeiro princípio, McCarthy & Prince (1986) adotam a hierarquia proposta por Selkirk (1982) para a organização dos constituintes prosódicos, de modo a poder atribuir à prosódia o fato de que certas operações morfológicas não se dão apenas pelo encadeamento de elementos morfológicos, como é o caso dos processos não concatenativos.

O segundo princípio requer que o molde satisfaça condições de boa-formação prosódica universais ou específicas de cada língua. Isso significa que o conteúdo circunscrito poderá ser posteriormente filtrado por um conjunto de condições de boa-formação, como, por exemplo a obrigatoriedade do *onset* e o licenciamento de determinados segmentos para a posição de coda silábica (GONÇALVES, 2013, p.352). Isso pode ser visto no esquema proposto por Gonçalves (2009), adaptado de Benua (1995):

(2)



A Morfologia Prosódica pode ser considerada um modelo transderivacional, pois a relação entre *input* e *output* não é direta, mas mediada pelo nível do molde. Gonçalves (2009, p.74) afirma que “por força da circunscrição prosódica, condições de minimalidade atuam no *input*, gerando um *output* (molde), que, por sua vez, passa a ser o *input* sobre o qual podem atuar determinadas condições de boa-formação (sobretudo silábicas e fonotáticas).” Quando essas condições de boa-formação são satisfeitas, o *output* final é formado.

O domínio de que fala o princípio (iii) é o domínio da base sobre o qual se dará o processo morfológico. Os processos morfológicos têm sempre como foco determinado domínio

da base. Em um processo como a derivação sufixal, por exemplo, os sufixos podem se adjungir a temas (**‘comunicação’**, **‘falsificação’**) ou a raízes (**‘timeco’**, **‘livreco’**). Nesse caso, o domínio da base selecionado pelo processo apresenta motivação puramente morfológica.

Entretanto, Belchor (2014), em seu estudo sobre o truncamento, observa que a morfologia não concatenativa

“é caracterizada pelo fato de as informações morfológicas não se sucederem linearmente – daí a inviabilidade de analisar um processo de redução tal como o truncamento de forma que se busque, na sequência apagada, um elemento de especificação estritamente morfológica, assim como um afixo. Por essa razão, o fenômeno deve ser analisado de modo que o foco seja estabelecido na porção que efetivamente constituirá o truncamento, uma vez que esta se mostra altamente regular.” (p.96)

Portanto, no caso de processos como o truncamento — e que também pode se aplicar à hipocorização, como mostra Gonçalves (2004b) — o domínio da base sobre o qual atua o processo morfológico pode ser selecionado não apenas por motivações morfológicas, mas também por motivações prosódicas. Por isso o papel da circunscrição prosódica é fundamental para essa teoria. É devido ao caráter prosódico identificado nos processos não concatenativos que boa parte dos pesquisadores que analisam esses processos fazem uso da Morfologia Prosódica em suas análises. Diversos autores analisaram o truncamento a partir da Morfologia Prosódica, ou utilizam os princípios da MP com base nos instrumentos da Teoria da Otimidade, como é o caso da Teoria da Correspondência.

2.2.1.1 Araújo (2002)

Araújo (2002) analisa o fenômeno do truncamento e da reduplicação. O autor caracteriza esses fenômenos como marginais, porém, produtivos no português e os denomina de não lineares ou não concatenativos, pois não há o encadeamento de morfemas na sua formação. No truncamento, o encadeamento é substituído por supressão de elementos na palavra. Em outros processos não-lineares há também a cópia total ou parcial da base (reduplicação), sobreposição ou amálgama (*blending* ou *portmanteaux*), acronímia (siglas), entre outros.

De acordo com o autor, o truncamento se caracteriza pela “redução de uma palavra-matriz sem perda de valor semântico” (ARAÚJO, 2002, p. 64). Essa redução gera a perda de material segmental no limite direito da palavra e, para que o truncamento ocorra, a palavra-matriz deve ter, no mínimo, três sílabas. Os resultados encontrados por Heineck (2018) vão de encontro a essa proposta de Araújo (2002), mostrando que pode haver perda de material

segmental também no limite esquerdo da palavra (paranoia > noia) e que a palavra-matriz pode possuir duas sílabas (chinês > china).

Outra característica importante do truncamento, destacada pelo autor, é a de que a forma truncada e a palavra-matriz devem ser semântica e pragmaticamente intercambiáveis, ou seja, que não há perda ou mudança de sentido na troca de uma forma pela outra. Alguns autores, como Gonçalves (2004) e Martini (2010), enxergam mudanças pragmáticas em relação ao truncamento e sua palavra-base. Martini (2010, p. 183) enfatiza que a palavra-matriz e a palavra truncada são intercambiáveis semanticamente, ou seja, dentro de uma sentença essas palavras são sinônimas de conteúdo. Entretanto, não são intercambiáveis pragmaticamente.

Levando em consideração as características definidas pelo autor, Araújo classifica alguns dados como *pseudotruncamentos*. É o caso da forma *deprê*, que poderia ser derivada de *depressão*, *deprimido(a)* ou *depressivo(a)*. Nesse caso, portanto, o falante seria incapaz de rastrear a palavra matriz que gerou a forma *deprê*, pois essa forma permite que mais de uma palavra-matriz seja a base.

O autor considera, ainda, como prováveis pseudotruncamentos, as formas derivadas de afixos cultos de origem greco-latina, germânica, etc., ou de palavras com estrutura semelhante à composição por afixos.

(3) minissaia, minidicionário, etc.	míni
microcomputador, microscópio, etc.	micro
fotografia, fotocélula, etc.	fóto
gastrologia, gastrocentro, etc.	gástro
eletrocardiograma, eletrodoméstico, etc.	elétro ¹

(ARAÚJO, 2002, p. 66)

Nos casos apresentados nos exemplos acima, a palavra-matriz é recuperável contextualmente, segundo o autor. Em um exemplo como *eletro*, a palavra de origem é identificável no contexto de uso e a forma truncada é usada metonimicamente. Essa falta de delimitação precisa entre o truncamento e os pseudotruncamentos tem gerado análises equivocadas do fenômeno, segundo Araújo. Por esse motivo, truncamentos têm sido tratados na literatura como aleatórios e imprevisíveis.

Os exemplos classificados como pseudotruncamentos por Araújo são incluídos nas análises de outros autores como truncamentos legítimos. No caso de *deprê*, o contexto de uso

¹ A notação aqui utilizada para demarcar o acento do truncamento é a notação do próprio autor (ARAÚJO, 2002), que faz uso do acento gráfico para demarcar a sílaba tônica da forma truncada (ex. *refrí*, *deprê*).

da forma torna possível o resgate da palavra de origem, como podemos ver nos exemplos propostos abaixo:

- (4) A *deprê* é um problema sério da sociedade / A depressão é um problema sério da sociedade
Ele tem andado muito *deprê* / Ele tem andado muito deprimido
Ela é uma pessoa *deprê* / Ela é uma pessoa depressiva

Acreditamos, portanto, que esses casos não devem ser classificados como *pseudotruncamentos*. No que diz respeito às formas derivadas de afixos cultos de origem greco-latina, germânica, etc., tomamos como base as propostas de Gonçalves e Vasquez (2004) e Scher (2011), que incluem essas formas nas suas classificações de tipos de truncamento.

Neste trabalho, o autor propõe uma descrição do truncamento que evidencie a previsibilidade e regularidade do processo, mostrando que o truncamento obedece a mecanismos possíveis de formação de palavras da língua. Como citado anteriormente, Araújo percebe regularidade no processo no fato de que há sempre uma redução de uma palavra-matriz (com no mínimo três sílabas) para uma palavra truncada, que geralmente é formada por duas sílabas. Essa redução privilegia o limite esquerdo da palavra-matriz. A posição do acento no truncamento é variada e exige uma análise mais cuidadosa. Em alguns casos o acento recai na penúltima sílaba (5a), em outros na última sílaba (5b).

- (5) a. *cerveja* → *cérva*
 neurose → *nêura*
 flagrante → *flágra*
b. *professor* → *prófi*
 bijuteria → *bijú*
 refrigerante → *refrí*

(ARAÚJO, 2002, p. 68)

No grupo (5a) estão palavras-matriz com três sílabas e acento na penúltima sílaba, o que implica que não possuem acento secundário. Nesses casos, formam-se, invariavelmente, formas truncadas com duas sílabas e paroxítonas. No grupo (5b) estão as palavras com três ou mais sílabas e acento secundário. Nesses casos, o acento da forma truncada recairá na sílaba que carregava o acento secundário da palavra-matriz. O autor afirma, no entanto, que é possível que

haja variação dialetal no acento secundário, o que implica em diferenças na acentuação das formas truncadas. O autor acredita que isso se deva à diferença da posição do acento secundário na palavra *exposição* em cada um dos dialetos (*èxposição* no dialeto paulista; *expòsição* no dialeto paraense).

A partir da análise e descrição de seus dados, Araújo (2002, p. 70) propõe condições para que o truncamento ocorra:

- (6) (i) a palavra-matriz deve ter três ou mais sílabas;
- (ii) contando-se da esquerda para a direita, selecionam-se as duas primeiras sílabas da palavra e eliminam-se as restantes: $\sigma_1 \sigma_2 \sigma_3 \sigma_4 \sigma_5 \sigma_6 \dots \rightarrow \sigma_1 \sigma_2$;
- (iii) as condições fonotáticas da língua devem ser respeitadas;
- (iv) o acento deve ser atribuído à sílaba que guardava acento secundário na palavra-matriz. Se não houver acento secundário, acentua-se a penúltima sílaba.

Araújo também cita em sua análise os truncamentos trissílabos, como em *delegado* > *delega*, *português* > *portuga*, embora os considere menos comuns que os dissílabos. De acordo com o autor, essas formas são paroxítonas e tendem a preservar a raiz da palavra-matriz de modo a manter a lexicalidade da palavra. O autor destaca que há sempre o acréscimo de uma vogal -a nessa classe de truncamentos e classifica essa vogal como uma epêntese, que teria a função de impedir formações com codas finais proibidas (*português* > **portug*).

O autor conclui sua análise do truncamento identificando uma tendência geral para o processo: formação de uma palavra dissilábica com sílaba final aberta e a manutenção de estruturas prosódicas alinhadas à direita na palavra-base. Os resultados do nosso estudo (HEINECK, 2018) corroboram, de um modo geral, a tendência encontrada por Araújo. No entanto, algumas das condições para a ocorrência do truncamento propostas pelo autor não se confirmam nos nossos dados, como é o caso da exigência de uma palavra-base com, no mínimo, três sílabas. Araújo trabalhou apenas com poucos dados do dialeto paulista e mineiro², o que pode ter limitado um pouco a análise. Sabemos que a coleta de dados de truncamento não é tão simples, pois, apesar de ser um processo produtivo no português, não é tão frequente quanto outros processos. Um dos principais méritos do trabalho feito por Araújo está no fato de ser um dos primeiros a trabalhar com truncamento no português brasileiro.

² A lista de dados utilizados na análise é apresentada nos anexos do artigo.

2.2.1.2 Gonçalves (2004)

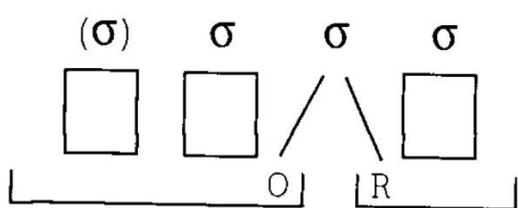
Gonçalves (2004) realiza um estudo bastante amplo de processos não concatenativos do português e tem como principal objetivo descrevê-los de forma sistemática, procurando estabelecer a regularidade dos processos. De acordo com o autor, os processos não concatenativos tem como função ampliar o vocabulário e expressar carga emocional variada (GONÇALVES, 2004, p. 9). Os processos não concatenativos foram vistos por muitos autores como malcomportados por não se encaixarem nos modelos aglutinativos da gramática, como é o caso de Item-e-Arranjo. Os processos não concatenativos são constituídos de uma morfologia que não é pura, mas requer acesso a informações prosódicas.

Gonçalves divide os processos não concatenativos em três tipos: afixação não linear, encurtamento e fusão. A afixação não linear é composta pelo processo de reduplicação (Fátima - Fafá; Puxa-Puxa). No encurtamento estão os processos de truncamento (salafrário - salaфра) e hipocorização (Fernando - Nando). A fusão engloba os processos de siglagem (PT – Partido dos Trabalhadores) e mesclagem lexical (Chá + Café = Chafé).

Em relação ao truncamento, o autor afirma que é um recurso morfológico de natureza expressiva, que não apresenta função lexical, mas que “revela o ponto de vista do falante sobre o que diz, chamando a atenção do seu interlocutor para algo avaliado negativamente”. Em relação ao formato morfoprosódico do truncamento, Gonçalves define o processo como simultaneamente não concatenativo (cópia de parte da palavra-base) e aglutinativo (acréscimo da vogal final). Diferentemente do que aponta Araújo (2002), Gonçalves afirma que o truncamento não leva em conta o acento lexical da palavra-base e sempre forma paroxítonas. Em Heineck (2018), entretanto, apesar da tendência de formas truncadas paroxítonas, também encontramos dados de truncamentos oxítonos.

Com uma análise que faz uso da Morfologia Prosódica, Gonçalves (2004) propõe que, no truncamento, a circunscrição mapeia uma sequência que não aparece na forma diminuta, ou seja, a circunscrição é negativa (p. 20). A sequência mapeada é descartada e acrescenta-se a vogal -a à sequência restante.

abaixo:



Forma de Base

Desassociação + -a

← circunscrição negativa

parâmetro do pé: forme um pé binário [σ σ]

parâmetro da direcionalidade: E ← D # #

parâmetro do alinhamento: alinhe -a à direita da Forma de Base, após desassociação

Ex.: ba.	te.	r (is. ta)	+ -a = batera
ves.	ti.	b (u. lar)	+ -a = vestiba
	fla.	gr(na. te)	+ -a = flagra
	sa.	p (a. tão)	+ -a = sapa

(GONÇALVES, 2004, p.20)

Na representação vê-se que a circunscrição mapeia um pé binário da direita para esquerda, deixando apenas o *onset* da penúltima sílaba, com o qual se adjunge a vogal -a. A parte mapeada é descartada e o que é preservado é o material restante à esquerda do material descartado.

Segundo Gonçalves, a hipocorização se sujeita a Condições de Boa-formação Silábica, que privilegia, na formação do hipocorístico, estruturas silábicas menos complexas, como vê-se em *Alexandre > Xande*, em que o *onset* complexo da palavra-matriz não permanece no hipocorístico. O truncamento, pelo contrário, não segue necessariamente essas condições, não impede a presença de *onsets* complexos, por exemplo. Por esse motivo, o autor afirma que o Truncamento privilegia a Fidelidade sobre a Marcação, ou seja, é mais fiel à base, preservando uma sequência da palavra-matriz.

É importante observar que Gonçalves realiza apenas a análise de formas truncadas trissílabas com acréscimo da vogal -a. Isso pode ser bastante restritivo para a análise do fenômeno, visto que vários autores identificam dados e padrões diferentes de truncamentos que não são abordados nessa análise. As formas truncadas como *noia* (paranoia) e *mor* (amor), por exemplo, não se encaixariam na análise proposta pelo autor, pois, se a circunscrição é negativa e mapeia um pé binário da direita para a esquerda, a margem da palavra-base a ser excluída seria justamente a que forma o truncamento.

2.2.1.3 Gonçalves e Vasquez (2004)

O estudo de Gonçalves e Vasquez (2004) sobre o truncamento tem como base teórica a Teoria da Otimidade (OT), mais especificamente a Teoria da Correspondência, que traz uma interface da fonologia com a morfologia, um dos princípios da Morfologia Prosódica. A Teoria da Correspondência trata não somente da relação de fidelidade do tipo *input-output* (I-O), mas também da relação entre *outputs* (O-O), permitindo, dessa forma, de acordo com Gonçalves (2009, p.217), diferenciar operações morfológicas de processos puramente fonológicos, pois “as primeiras podem atuar numa dimensão de correspondência O-O, enquanto os últimos se

restringem à dimensão I-O”. O principal objetivo dos autores é estabelecer uma hierarquia entre as restrições que estão relacionadas ao fenômeno do truncamento.

Os autores classificam os dados de truncamento em quatro tipos, de acordo com suas especificidades:

- (i) TRUNC + A: preserva a base da palavra e se acrescenta uma vogal especificada do truncamento /-a/. Exemplos: cafajeste > cafa; Barcelona > Barça
- (ii) Palavras cortadas preservam a vogal que está presente na sua estrutura prosódica. Exemplos: depressão > deprê; refrigerante > refri
- (iii) Corte se dá em uma consoante:
 - a. Com epêntese: corte em uma consoante oclusiva. Exemplo: trabalho > trab(i)
 - b. Sem epêntese: corte em uma consoante licenciada para coda. Exemplo: aniversário > aniver
- (iv) Corte na base à esquerda preservando-se o morfema inicial, que pode ser um radical preso ou um prefixo. Exemplo: psicologia > psico; odontologia > odonto; pentacampeão > penta.

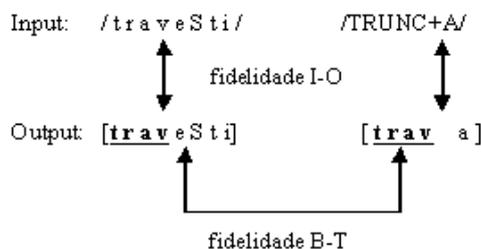
De modo a detectar padrões de uso gerais e variáveis para as formas truncadas, os autores desenvolveram e aplicaram um teste composto de quatro questões. Na primeira questão, os participantes precisavam descrever o significado de palavras truncadas que, teoricamente, eram de uso generalizado e, portanto, conhecidas pelos participantes. Na segunda questão, eram apresentados a palavras polissílabas e deveriam encurtá-las. A terceira questão se assemelha à primeira, com a diferença de que os dados de truncamentos são desconhecidos pelos participantes, pois foram criados pelos pesquisadores. Na quarta e última questão, os participantes precisavam marcar os contextos de uso dos truncamentos. Eles eram apresentados a treze situações que variavam de um grau máximo de formalidade até um grau mínimo.

A partir dos resultados desse teste, os autores propuseram uma hierarquia genérica para o fenômeno do truncamento, na qual as restrições de fidelidade aparecem em dois lugares diferentes da hierarquia.

FIDELIDADE > MARCAÇÃO > FIDELIDADE

De acordo com Gonçalves e Vasquez (*op. cit*), as restrições de fidelidade são de extrema relevância para o truncamento, pois “são elas que vão propiciar ao falante a recuperação da base de uma palavra truncada” (p.61). No esquema a seguir, visualiza-se como se dá essa dupla associação de fidelidade.

Figura 1 - Esquema de hierarquia de restrições de fidelidade e marcação no truncamento.



Fonte: Gonçalves e Vasquez (2004, p. 61)

Como se vê no esquema, na posição hierárquica mais alta estão as restrições de fidelidade I-O (*input-output*), e, na mais baixa, as de fidelidade O-O (*output-output*). No primeiro nível de fidelidade, temos, no *input*, somente a vogal /-a/ como elemento plenamente especificado. O morfema /TRUNC/ não possui material fônico, é um morfema vazio. A restrição MAX-IO é violada, portanto, quando a vogal /a/ não estiver presente no *output*. O segundo nível de fidelidade é base-truncamento. Neste nível, a restrição MAX-BT será violada toda vez que algum segmento da base for apagado. Os autores propõem a seguinte hierarquia para as restrições:

MAX-IO >> ALIN-B, esq, T, esq >> TODO-PÉ (D) >> TROQ >> ONSET >> MAX-BT >> ANAL-s

De acordo com os autores, a restrição MAX-BT, apesar de ocupar uma posição baixa na hierarquia, será, quase sempre, responsável pela escolha do candidato ótimo, que será aquele que violar menos essa restrição. Isto confirma que o truncamento é um processo que prioriza fidelidade em detrimento de marcação, pois é importante que o falante possa recuperar a base de uma forma truncada.

2.2.1.4 Martini (2010)

Martini (2010) propõe uma análise que integra os processos de truncamento e hipocorização. A autora acredita que o truncamento deve ser compreendido como um subgrupo da hipocorização. De forma semelhante a Araújo (2002), a autora também não considera, em sua análise, formas “que se estruturam com a manutenção integral do morfema à esquerda da forma de base” (p.254), como, por exemplo, *vídeo* < *videocassete*, *odonto* < *odontologia*, *eletro* < *eletrocardiograma*, pois os classifica como processos puramente morfológicos. O processo

do truncamento é dividido em cinco padrões e procura-se estabelecer hierarquias de restrições que expliquem a formação de cada um deles.

O primeiro padrão (Tipo-A) compreende formas que preservam a sílaba mais à esquerda da forma de base e, por isso, a sua formação faz referência apenas à estrutura silábica; a estrutura prosódica da palavra-base não é tão importante. As sílabas selecionadas para a manifestação dos Truncamentos monossilábicos ancorados à esquerda são as sílabas iniciais da forma de base que possuem *onset*. Desse modo, aqueles truncamentos cujas sílabas iniciais da palavra-matriz correspondente estejam sem onset não formarão truncamentos monossilábicos (MARTINI, 2010, p. 191-192). Alguns exemplos desse padrão são *bi* < *bilhão*, *Fla* < *Flamengo*, *Lu* < *Luciana*.

O segundo padrão, denominado como Tipo-B, diz respeito a formas que preservam o pé principal da forma de base. O corte ocorre à esquerda da forma de base, como nos exemplos *mor* < *amor*, *Mengo* < *Flamengo*, *Nense* < *Fluminense*. Neste padrão, a circunscrição positiva mapeia, da esquerda para a direita, o pé principal da palavra-matriz -matriz, cujo acento poderá recair sobre a última sílaba (cf. *ssór* < *professor*), favorecendo formas oxítonas, ou sobre a penúltima sílaba (cf. *Méngo* < *Flamengo*, *cháça* < *caçaça*), favorecendo formas paroxítonas. (MARTINI, 2010, p. 192)

O terceiro padrão (Tipo-C) engloba formas truncadas com cópia integral das duas primeiras sílabas da palavra-matriz e acento paroxítono. Os truncamentos deste tipo não levam em conta a pauta acentual da forma de base. A autora afirma que as formas truncadas deste tipo “sempre apresentam duas sílabas e que tendem a manter a complexidade silábica da forma de base (cf. *prófe* < *professor*, *transa* < *transação*) e a manutenção da coda existente na respectiva palavra-matriz (cf. *cárna* < *carnaval*, *Pínda* < *Pindamonhangaba*, *transa* < *transação*)” (MARTINI, 2010, p. 195). São exemplos de formas truncadas do Tipo-C *palha* < *palhaço*, *bisa* < *bisavó*, *profe* < *professor*.

Ao quarto padrão, denominado Tipo-D, são aplicadas as mesmas restrições do Tipo-C e, portanto, ele possui as mesmas características do Tipo-C, com exceção do acento, visto que os truncamentos do Tipo-D são sempre oxítonos. São exemplos de formas truncadas do Tipo-D *belê* < *beleza*, *expô* < *exposição*, *refri* < *refrigerante*.

O último padrão (Tipo-E) se diferencia mais dos outros padrões. As formas truncadas podem ser dissilábicas ou trissilábicas e há sempre a inserção de uma vogal final ou sílaba CV que não estava presente na forma de base, como nos exemplos *sarja* < *sargento*, *motora* <

motorista, sanduba < *sanduíche*. Em decorrência disso, de acordo com a autora, foi necessária a inserção da restrição RM, abaixo definida, na hierarquia, como uma restrição dominante.

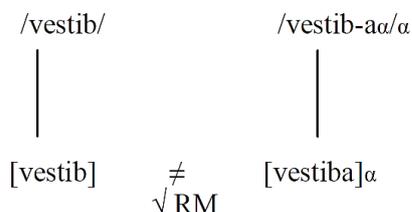
(7) *Realize Morpheme (RM)*

Dado que α seja uma forma morfológica, β seja uma categoria morfossintática e $F(\alpha)$ seja a forma fonológica da qual $F(\alpha + \beta)$ é derivada para expressar uma categoria morfossintática β . Então, RM será satisfeita em relação à β se, e somente se, $F(\alpha + \beta) \neq F(\alpha)$ fonologicamente.

(MARTINI, 2010, p. 238)

Portanto, RM é satisfeita se o candidato for não idêntico ao *output* da raiz, enquanto a restrição é violada se ambas as formas compartilharem exatamente da mesma representação fonológica. A não identidade requerida para a satisfação de RM não está limitada por nenhum aspecto fonológico específico da base (MARTINI, 2010, p. 241). Nas formas truncadas do Tipo-E, é necessário inserir RM no ranking de restrições, pois há a inserção de material fonológico à base, que, nesse caso, é o sufixo *-a*, como vemos na figura abaixo.

Figura 2 - Representação da inserção de material fonológico à base do truncamento.



Fonte: Martini (2010, p. 244)

Em um truncamento do Tipo-E, RM é plenamente satisfeita porque, em todo candidato desse tipo, o *output* da raiz não contém nenhum afixo, mas o *output* da forma derivada carrega um segmento sufixal.

2.2.1.5 Belchor (2014)

Belchor (2014) faz uma análise do truncamento com base nos pressupostos da Morfologia Prosódica Circunscritiva. Os dados utilizados na análise são de um *corpus* coletado, em grande parte, por Gonçalves e Vazquez (2004) e complementado por Belchor (2009). Os dados foram coletados de jornais de grande circulação, como *O Globo* e *Jornal do Brasil*; dicionários eletrônicos, como Aurélio e o dicionário informal; além de dados produzidos em diversas situações de interação entre falantes. Os dados foram divididos em três padrões: ‘refri’,

‘odonto’ e ‘flagra’ e cada um deles é analisado separadamente de acordo com a MP Circunscritiva.

O padrão ‘refri’ engloba as formas truncadas que apresentam estrutura composta pelas duas primeiras sílabas da base e formam um pé iâmbico. Nos termos da MP Circunscritiva, de acordo com a autora, nesses dados, a circunscrição “mapeia as duas primeiras sílabas da palavra-matriz e envia essa informação para o molde, formatado para o posicionamento da cabeça do pé à direita — o que garante a formação de pés iâmbicos e, por conseguinte, de *outputs* oxítonos.” (p. 103).

A circunscrição é positiva nos dados do padrão ‘refri’, ou seja, a porção mapeada é a enviada para o molde. O mapeamento ocorre da esquerda para direita, a partir do início da palavra-matriz, e a instrução determina que sejam circunscritas as duas primeiras sílabas integrais da base. O formato do molde exige que a forma truncada apresente a cabeça do pé à direita, formando, portanto, um pé iâmbico.

No padrão ‘odonto’, a base das palavras truncadas é, em sua maioria, de compostos formados por dois ou mais radicais de origem grega ou latina, como em *fonoaudiologia* e *eletrocardiograma*, ou de um radical adjungido a uma palavra, como em *heterossexual* e *microcomputador*. Há ainda dados que se caracterizam pela presença de um prefixo, como em *pré-vestibular* e *ex-namorado*. Os truncamentos do padrão ‘odonto’ caracterizam-se pela preservação integral do radical situado mais à esquerda da palavra-matriz. Em relação ao formato prosódico, a autora afirma que o molde consiste em um pé trocaico e, por essa razão,

a circunscrição mapeia o primeiro morfema a partir da margem esquerda da base e envia essa informação para o molde, configurando melodicamente como um troqueu moraico do tipo (* .) no caso dos dados que possuem duas ou mais sílabas (‘últra’; ‘elétro’), ou do tipo (*), no caso dos monossílabos (‘êx’; ‘pós’). (BELCHOR, 2014, p. 112)

O terceiro padrão, ‘flagra’, pode ser considerado misto, visto que há um equilíbrio entre a relevância de fatores prosódicos e morfológicos, pois o mapeamento da base delimita o radical, priorizando a informação morfológica na circunscrição; porém, o material circunscrito deve atender a condições prosódicas para formar o *output*. A tonicidade dos *outputs* é, de forma categórica, paroxítona, formando pés trocaicos. Os radicais mapeados são, em sua maioria, terminados em consoantes que não podem ocupar a posição final de um vocábulo, dessa forma, fazendo-se necessário o acréscimo da vogal -a.

Diferentemente de Gonçalves (2004), Belchor propõe que a circunscrição no padrão ‘flagra’ é positiva e ela ocorre da esquerda para a direita, a partir do início da base, como nos outros padrões. A circunscrição mapeia um único morfema da base, no caso, o radical. No entanto, segundo a autora, há alguns casos nos quais o radical não é integralmente aproveitado no *output* e, por isso, devem ser observadas algumas condições que atuam sobre o material resultante da circunscrição. A primeira condição exige que o molde deva ser terminado em consoante para que haja a afixação do marcador -a. Isso faz com que, no caso de radicais terminados em vogal, ela seja suprimida. A segunda condição impõe um limite de três sílabas para o *output*. A terceira condição também tem relação com o tamanho da forma truncada e exige que a parte da base enviada para o molde seja uma sequência que, acrescida do marcador -a, não se torne igual nem maior que a base³.

2.2.2 Morfologia Distribuída (DM)

A Morfologia Distribuída (DM), proposta por Halle & Marantz (1993), foi desenvolvida a partir da Teoria Gerativa e é uma teoria morfossintática de gramática. Neste modelo de gramática, palavras e sentenças são formados por um único componente gerativo da gramática: a sintaxe. Não há nenhum componente gerativo lexical, por esse motivo, tudo que previamente havia sido atribuído ao léxico em modelos de gramática anteriores precisa, no modelo da DM, ser discutido em termos de propriedades formais, semânticas e fonológicas, que estão distribuídas nos diferentes componentes da gramática (SCHER, 2016, p. 338).

São três as principais propriedades que definem essa teoria:

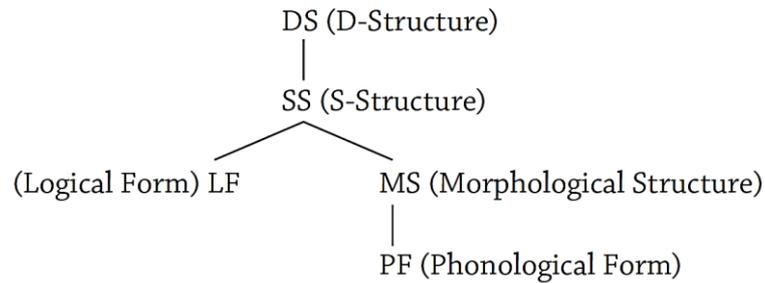
- (i) **Inserção lexical tardia** – as categorias sintáticas são puramente abstratas, sem nenhum conteúdo fonológico. A expressão fonológica dos nós sintáticos é inserida somente no mapeamento da Forma Fonológica (PF).
- (ii) **Subespecificação** – expressões fonológicas não precisam ser plenamente especificadas para as posições sintáticas nas quais serão inseridas. Itens de vocabulário *default* podem ser inseridos quando não há um item mais especificado disponível.

³ Aqui, cabe questionar o que a autora entende como “menor ou igual à base”. Se o tamanho for medido em número de sílabas, há contraexemplos para essa condição proposta por Belchor. Há exemplos de truncamentos como *português* > *portuga*, *chinês* > *china* e *sapatão* > *sapata*, nos quais o número de sílabas da forma truncada se mantém o mesmo da sua base.

- (iii) **Estrutura hierárquica *all the way down*** – há apenas um único sistema gerativo responsável tanto pela estrutura morfológica quanto pela estrutura sintática.

Levando em conta essas propriedades, a arquitetura da gramática é a representada no esquema abaixo.

(8)



Fonte: Halle & Marantz (1993)

De acordo com Scher (2016, p. 339), não é incomum haver divergências entre a organização dos nós terminais no nível sintático, de um lado, e na Estrutura Morfológica (MS) e na Forma Fonológica (PF), de outro. Por esse motivo, a DM prevê que algumas operações morfológicas irão se aplicar a estruturas sintáticas antes de elas serem preenchidas em termos de expressão fonológica. Nesse sentido, nós terminais podem ser adjungidos à estrutura, podem se mover ou ainda ter alguns dos traços trazidos da sintaxe deletados em alguns contextos. É nesses casos que processos como o truncamento, estudado pela autora, podem se encaixar.

2.2.2.1 Scher (2011)

Scher (2011) analisa dados de truncamento do português brasileiro e do espanhol peninsular sob a perspectiva da Morfologia Distribuída (DM). A partir desta perspectiva, não há a presença de um léxico gerativo na arquitetura da gramática e processos de formação de palavras (incluindo o truncamento) se dão no componente sintático da gramática. A autora propõe sete padrões para o truncamento que se aplicam nas duas línguas analisadas, com pequenos ajustes.

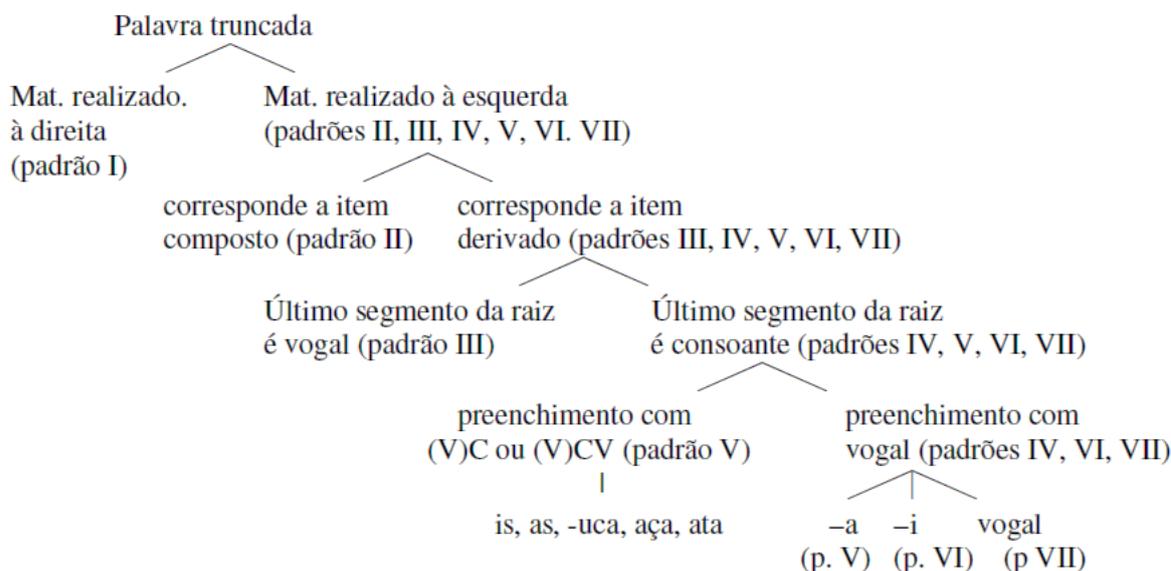
- (i) **Padrão I:** a forma truncada corresponde a formas composicionais ou derivacionais e realiza apenas sua parte final. (*noia* < *paranoia*; *fessô* < *professor*; *chaça* < *cachaça*).

- (ii) **Padrão II:** forma truncada corresponde a um composto e realiza apenas o morfema inicial (*psico* < *psicologia*; *super* < *supermercado*; *eco* < *ecografia*).
- (iii) **Padrão III:** forma truncada corresponde a uma palavra bimorfêmica⁴ e realiza parte da raiz de sua forma correspondente, mantendo uma vogal que está presente na sua estrutura silábica (*deprê* < *depressão*; *preju* < *prejuízo*; *pregui* < *preguiça*).
- (iv) **Padrão IV:** forma truncada corresponde a uma palavra bimorfêmica, ou interpretada como tal pelo falante, e preserva a raiz da palavra correspondente, acrescentando-se a ela a vogal -a (tipo mais frequente em PB) (*salaфра* < *salafrário*; *reaça* < *reacionário*; *delega* < *delegado*).
- (v) **Padrão V:** forma truncada corresponde a uma palavra bimorfêmica, ou interpretada como tal pelo falante: raiz, ou parte da raiz da palavra correspondente + sequência (V)C *-(a)s* ou *-(i)s* ou (V)CV *-(u)ca* (*migs* < *amigo*; *mamis* < *mamãe/mãe*; *legas* < *legal*; *feijoca* < *feijoada*).
- (vi) **Padrão VI:** forma truncada preserva a consoante final da raiz. É dividida em dois grupos: com epêntese (consoante final oclusiva: *trab(i)* < *trabalho*; *bob(i)* < *bobeira*; *net(i)* < *internet*); e sem epêntese (consoante final licenciada para coda: *mongol* < *mongoloide*; *Fortal* < *(Fortaleza)*; *niver* < *aniversário*).
- (vii) **Padrão VII:** a forma truncada corresponde a uma forma derivada. Preserva a raiz até sua consoante final e insere uma vogal, mesmo nos casos em que não há proibição para a coda da sílaba final. (*envase* < *envasamento*; *encaixe* < *encaixamento*; *escape* < *escapamento*; *fabrico* < *fabricação*).

A observação dos padrões do truncamento levou a autora a formular novos questionamentos. A autora sugere uma reorganização dos padrões que leva em conta fatores como: (a) se o material que se realiza na forma truncada estava à direita ou à esquerda da representação em sua estrutura subjacente; (b) se a palavra-base era formada por composição ou derivação; (c) se o último segmento da raiz nas formas truncadas é uma vogal ou consoante e se, sendo uma consoante, há inserção de material fonológico; e (d) qual é esse material inserido. Como resposta a essas questões, Scher constrói o diagrama abaixo.

⁴ De acordo com a proposta de Scher (2011), aliado ao modelo proposto pela Morfologia Distribuída, compreende que a formação dos truncamentos se dá a partir da manutenção da raiz e não da supressão de parte da palavra. Esse resultado busca corroborar a premissa da DM de que raízes acategoriais, e não bases ou radicais, são o ponto de partida para a formação de palavras. É na derivação que essas raízes se associam a núcleos funcionais que atribuem categorias a elas e é nessa etapa que a palavra pode ser compreendida como bimorfêmica.

Figura 3 - Diagrama do processo do truncamento proposto por Scher (2011)



Fonte: Scher (2011, p. 73)

Devido a semelhanças entre os sete padrões sugeridos por Scher, a autora sugere que se reduzam os padrões das formas truncadas para três: os padrões I, II e um terceiro padrão que reúne os padrões de III a VII.

Diferentemente da perspectiva adotada por trabalhos anteriores que sugerem um forte condicionamento de fatores fonológicos na formação de truncamentos, Scher (2011) procura apontar a relevância de outras propriedades formais. A organização dos padrões proposta pela autora tem forte base morfológica e distribucional, pois leva em conta, por exemplo, “o tipo de processo morfológico que forma a palavra sinônima da forma truncada, e a posição (à direita ou à esquerda) do material realizado na forma truncada, em relação à sua estrutura subjacente” (SCHER, 2011, p. 73).

A análise dos dados e dos diferentes padrões do truncamento levam a autora a concluir que os padrões devem ser formulados com base em suas propriedades formais e que, entre as propriedades morfológicas, está o fato de que a raiz tem um papel importante na formação de formas truncadas, pois é sempre ela, ou parte importante dela, que permanece na forma truncada.

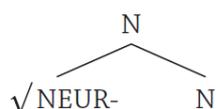
2.2.2.2 Scher (2016)

O trabalho realizado por Scher (2016) dá sequência ao seu estudo anterior (SCHER, 2011) e também tem como base teórica a DM. Nesta análise, a autora resgata uma proposta de análise concatenativa para o processo do truncamento, distanciando-se da grande maioria das análises anteriores do fenômeno, e tem como foco dois dos padrões propostos por Scher (2011): o Tipo III e o Tipo IV, que estão exemplificados em (8).

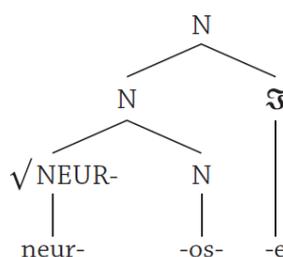
(9)	Tipo III	Tipo IV
	sarja – sargento	saudas – saudade
	neura - neurose	vont(i)s – vontade

Com base na estrutura de gramática proposta pela DM, a autora propõe que os truncamentos sejam tratados como derivação de uma raiz, e não como o resultado de um processo de apagamento de segmento da palavra-matriz. Scher (2011, p. 342) propõe que as formas truncadas são independentemente derivadas por um processo sintático que envolve categorização da raiz, seguido por uma operação morfológica que insere uma vogal temática na derivação. Na sequência, regras de inserção vocabular se aplicam e itens de vocabulário, como morfemas categoriais e sufixos temáticos são aplicados à raiz e a outros nós terminais. Esse processo é exemplificado nas representações em (10) e (11).

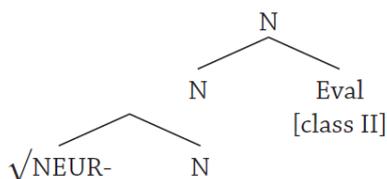
(10) a. syntactic derivation



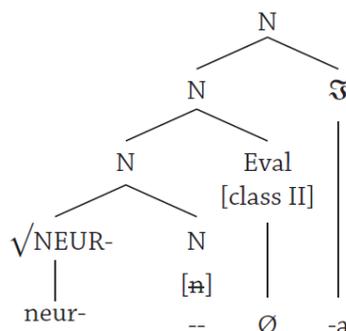
b. morphological derivation



(11) a. syntactic derivation



b. morphological derivation



A formação dos truncamentos de tipos III e IV envolve a presença de [EVAL], como vemos na representação acima. [EVAL] é uma categoria avaliativa que dá o caráter apreciativo às formas truncadas. Em (11b), vemos que o traço [n] é deletado. De acordo com a autora, antes da inserção vocabular, se aplica uma regra de empobrecimento que apaga o traço [n] na presença de [EVAL]. É importante destacar, no entanto, que, quando apagado, [n] já cumpriu seu papel de definir a classe gramatical da palavra que está sendo formada. [n] é apagado na inserção vocabular e, por isso, nenhum item de vocabulário será inserido na cabeça N. [EVAL] é preenchido por um item de vocabulário nulo e \mathfrak{N} é preenchido por morfemas de classe (ALCÂNTARA, 2010), como vemos em (12c).

- (12) a. [-os] → [n]
 b. \emptyset → [EVAL]
 c. [-o] → [classe I]
 [-a] → [classe II]
 [-e] → [classe III]

Em palavras monomorfêmicas como *cerveja* e *pijama*, para as quais poderia não parecer possível a formação de truncamentos, a autora propõe que uma reanálise ocorra para que os truncamentos *cerva* e *pija* surjam. Dentro dessa perspectiva, os falantes reanalisam palavras como *cerveja*, *pijama*, *fortuna* e *vestibular* como formas bimorfêmicas ou ainda multimorfêmicas (*cerv-*ej*-a*, *pj-am-a*, *fort-un-a* e *ves-tib-ul-ar*). Apesar de não serem considerados morfemas, as formas *-ej-*, *-am-*, *-un-*, *-ul-* e *-ar-* são vistas como tal pelos falantes. A evidência para isso estaria em outras palavras que também possuem essas formas, como *sertaneja* e *dinheirama*.

Por fim, a autora discute a qual das classes morfêmicas formais propostas por Alcântara (2010), as terminações *-as* e *-is*, que são acrescentadas a truncamentos como *bermas* e *tranquis*, corresponderiam. Iremos discutir essa proposta no próximo capítulo, quando trataremos de dados em que há o acréscimo de *-s*.

3 CONSTITUIÇÃO DO *CORPUS*

3.1 HEINECK (2018)

Em minha dissertação de mestrado (HEINECK, 2018), levantamos dados de truncamento e construímos um *corpus* organizado por critérios fonológicos e morfológicos da palavra-base e da forma truncada. Os dados do *corpus* foram usados por Heineck (2018) em dois experimentos que tinham como principal objetivo identificar possíveis padrões no processo do truncamento, bem como analisar o papel de elementos morfológicos e prosódicos no processo.

O objetivo mais geral do trabalho foi o de investigar e descrever o processo do truncamento, bem como compreender o seu funcionamento e as condições implicadas na sua ocorrência. No âmbito mais restrito, o trabalho se propôs a investigar o papel do acento e do número de sílabas da palavra-base na formação dos 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.

A metodologia usada no trabalho, bem com os principais resultados, serão descritos nas seções a seguir.

3.1.1 Levantamento de dados

Para a construção de um *corpus* de dados de truncamento, realizamos um levantamento dos dados reportados na literatura sobre o processo (ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014; SCHER, 2018). Além dos dados documentados na literatura, também foram registrados dados encontrados em redes sociais, programas de televisão ou produzidos em situações de interação entre falantes.

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 *online* e em um experimento presencial que eram compostos de um questionário social e três exercícios e foram realizados

no ano de 2018. O experimento *online* foi ancorado na plataforma SurveyMonkey, que permite a criação de questionários online para pesquisas de diversos fins. As respostas do questionário eram completamente anônimas, o nome do informante não era solicitado em nenhum momento e não tínhamos 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 *online*, esse experimento foi impresso em papel e os participantes respondiam às questões à mão. O experimento presencial contou com 46 participantes.⁵

A primeira etapa do experimento consistia em um questionário social. Os campos a serem preenchidos pelos informantes eram 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).

Na segunda etapa do experimento, os informantes eram 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 4.

Figura 4 – 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 _____

Fonte: Heineck (2018)

⁵ Os experimentos online e presencial completos estão nos anexos 10.8 e 10.9.

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 5. No experimento *online*, 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 5 – 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.

Fonte: Heineck (2018)

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. 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 6. Assim como ocorreu no exercício anterior, no experimento *online*, 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 6 – 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**.

Fonte: Heineck (2018)

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, em vez 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.

3.1.2 Principais resultados de Heineck (2018)

A partir da análise dos resultados dos experimentos, verificamos uma preferência por truncamentos dissílabos, paroxítonos e formados a partir das duas primeiras sílabas da palavra-base (*professor* – *profe*; *refrigerante* – *refri*). Houve também a ocorrência de truncamentos trissílabos em que há o acréscimo da vogal -a (*português* – *portuga*; *madrugada* – *madruga*). Nos casos de palavras com prefixos autoacentuados ou formados por duas raízes eruditas, a tendência foi de truncamentos formados pelo prefixo ou pela raiz mais à esquerda da palavra, como nos exemplos *pré-escola* – *pré* e *oftalmologista* – *oftalmo*.

No que diz respeito à relação do número de sílabas da palavra-base e do número de sílabas do truncamento, há uma tendência de a forma truncada ser menor do que sua base; no entanto, isso não pode ser entendido como regra, visto que também verificamos truncamentos com o mesmo número de sílabas da palavra-base. A tendência observada por Martini (2010,

p.236) em seus dados foi também observada nos nossos resultados: palavras dissílabas geralmente formam truncamentos também dissílabos, palavras trissílabas geralmente formam truncamentos dissílabos e palavras polissílabas tendem a formar truncamentos trissílabos.

Em relação à formação de diminutivos a partir de truncamentos, houve correspondência da vogal final do diminutivo com o gênero do truncamento na maioria dos casos. No entanto, em formas mais produtivas e frequentes da língua, como *foto* e *moto*, a maior parte dos informantes formou diminutivos terminados em *o*, a despeito do gênero dessas palavras. Podemos supor que, nesses casos, essas formas já façam parte do léxico dos falantes e, por isso, eles já as entendem como palavras femininas terminadas em *o*. Esse dado é interessante, pois há pouquíssimos dados de palavras femininas terminadas em *o* no português, e podemos supor que, em *foto* e *moto*, os falantes não identificam mais *o* como parte da raiz, mas como a vogal temática da palavra.

3.2 PAPEL DO NÚMERO DE SÍLABAS

3.2.1 O número de sílabas no truncamento

Entre os autores que realizaram estudos acerca do truncamento no português, não há muitos que analisam ou mencionam o papel do número de sílabas no processo. Araújo (2002) propõe uma série de condições que precisam se aplicar para que o truncamento ocorra. As condições propostas por Araújo tratam do número de sílabas tanto da base, quanto do produto. O autor afirma que a palavra-matriz deve ter, no mínimo, três sílabas e que a forma truncada será formada a partir das duas primeiras sílabas da palavra e, portanto, será dissílaba. O autor admite, ainda, que há alguns casos de truncamentos trissílabos, nos quais há o acréscimo de uma vogal epentética (*delegado*>*delega*). Esses casos seriam menos comuns do que os de truncamentos dissílabos, de acordo com o autor.

Vilela, Godoy e Silva (2006) não tratam especificamente do papel do número de sílabas da palavra-base na formação do truncamento. No entanto, contestam o trabalho de Araújo (2002) no que diz respeito aos truncamentos trissílabos. De acordo com dados coletados pelas autoras, o número de casos de truncamentos trissílabos é bastante expressivo e, portanto, esse tipo de formação deve ser considerado nas análises do fenômeno.

Dentre os estudos aqui levantados, o de Martini (2010) é o que apresenta uma análise mais sólida do papel do número de sílabas no processo do truncamento. A autora identifica, em sua análise, um padrão na formação do truncamento. De acordo com o seu estudo, em palavras-base dissílabas, como *chinês*, *burguês* e *mulher*, o truncamento também será dissílaba (*china*,

burga e mulha). Se a palavra-base for trissílaba, como em *japonês* e *flagrante*, a tendência será formar truncamentos dissílabos (*japa* e *flagra*, respectivamente). Já quando a palavra é polissílaba, o truncamento seria trissílabo, como em *baterista* > *batera* e *traficante* > *trafica*. Martini observa, em seus dados, que “formas derivantes mais curtas tendem a gerar truncamentos dissilábicos, ao passo que formas derivantes mais extensas tendem a formar truncamentos trissilábicos.” (p. 243).

No estudo realizado em minha dissertação de mestrado (HEINECK, 2018), como citado na seção anterior, identificamos uma tendência de o truncamento ser menor em número de sílabas em relação à sua palavra de origem. No entanto, também registramos truncamentos com o mesmo número de sílabas da base. Os nossos resultados seguiram a tendência identificada por Martini (2010). Alguns dados interessantes nesse sentido foram os truncamentos formados a partir de *futebol* e *juvenil*. Esperávamos truncamentos trissílabos com acréscimo de *a* – *futiba* e *juvena* – no entanto, a maior parte das respostas foram de *fute/futi* e *juve/juvi*, preferindo-se um truncamento dissílabo, seguindo a tendência proposta por Martini (2010).

Considerando o estudo de Martini (2010) e os resultados obtidos na dissertação de mestrado (HEINECK, 2018), temos como hipótese que o número de sílabas pode ter algum papel na formação de truncamentos, ou seja, que o número de sílabas da palavra base pode influenciar o tamanho de sua respectiva forma truncada.

3.2.2 Metodologia

Para a análise do papel da sílaba, utilizamos os dados obtidos nos exercícios 1 e 2 do experimento *online* desenvolvido em Heineck (2018) e que foram descritos no capítulo 3. Devido ao recorte desta etapa de trabalho, selecionamos somente os resultados referentes ao número de sílabas dos dados. Codificamos tanto o tamanho do dado apresentado ao informante na pergunta dos exercícios, quanto o dado da resposta do informante. Para a análise realizada no *software* SPSS, foram calculadas as médias de sílabas das respostas dos participantes para cada uma das condições em cada um dos exercícios analisados. Excluimos os dados de respostas em branco.

Em um primeiro momento, foram feitas análises descritivas dos dados para, então, partir para as análises estatísticas inferenciais com os testes adequados para o desenho da nossa amostra. Os resultados desses testes, as decisões de análises que foram tomadas, bem como a interpretação desses resultados, serão expostas na seção a seguir.

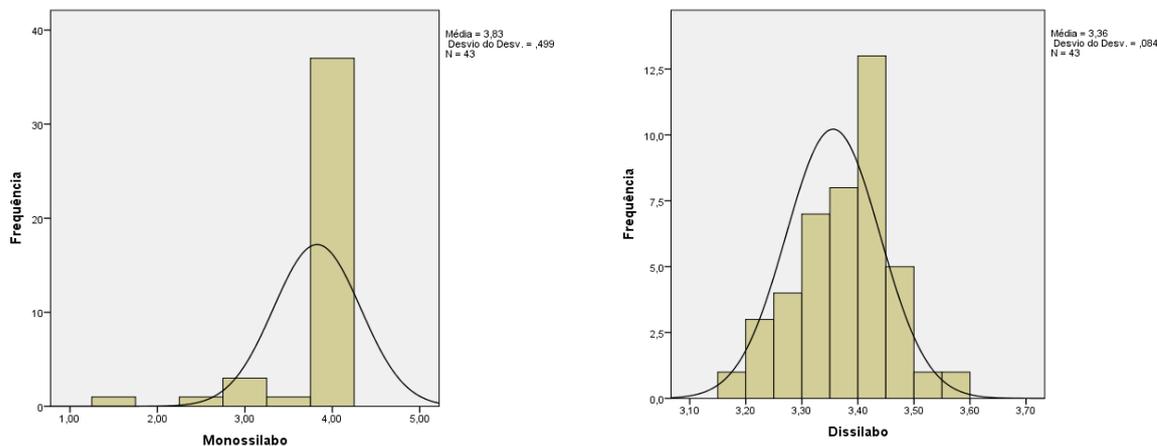
3.2.3 Resultados

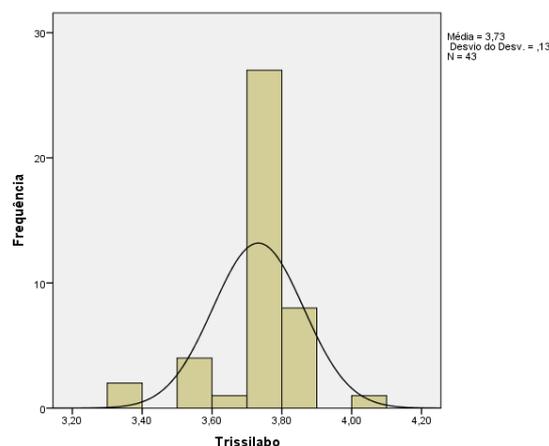
3.2.3.1 Exercício 1

No exercício 1 do experimento, os informantes eram apresentados a dados de formas truncadas e deveriam indicar a palavra-base para cada um dos truncamentos. Em relação ao número de sílabas, esse exercício tinha três condições: havia dados de truncamentos monossílabos, dissílabos e trissílabos. O objetivo, aqui, é o de verificar se essas condições influenciam nas respostas dos informantes, ou seja, se um truncamento dissílabo, por exemplo, terá como origem mais palavras-base trissílabas.

Tanto neste exercício quanto no exercício 2, todos os participantes responderam perguntas com dados de todas as condições. Portanto, o *design* do nosso experimento é intra-participantes, ou seja, o experimento tem os mesmos participantes em cada uma das condições das variáveis independentes. Para a escolha dos testes adequados para nossa análise, verificamos, primeiramente, se os dados se distribuíam normalmente. Os gráficos a seguir mostram a distribuição de cada uma das variáveis.

Gráfico 1 – Distribuição dos dados das variáveis *monossílabo*, *dissílabo* e *trissílabo* do exercício 1





Fonte: autor (2023)

Os gráficos parecem apontar para o fato de que não há uma distribuição normal em todas as variáveis. No gráfico de *monossílabo*, há uma grande concentração de dados em 4 (palavras-base com quatro sílabas), ou seja, quando eram apresentados truncamentos monossílabos, grande parte dos participantes propôs uma palavra-base polissílaba, e isso torna a nossa distribuição assimétrica positivamente. No gráfico da variável *trissílabo*, os dados também não acompanham a curva de normalidade. Apenas a variável *dissílabo* parece ter uma distribuição mais próxima de normal.

Para confirmar ou refutar a normalidade dos nossos dados, realizamos um teste Shapiro-Wilk de normalidade. Sendo a hipótese nula do teste Shapiro-Wilk a de que a população se distribui normalmente, os resultados das variáveis *monossílabo* ($p < 0,001$) e *trissílabo* ($p < 0,001$) evidenciam que seus dados não apresentam distribuição normal, pois refutam essa hipótese. Apenas os dados da variável *dissílabo* ($p = 0,179$) apresentaram distribuição normal. Tendo isso em vista, para poder comparar as variáveis entre si e verificar se há alguma diferença significativa entre elas, utilizamos um teste não-paramétrico equivalente a ANOVA de medidas repetidas, o teste Friedman.

Antes do teste Friedman, realizamos testes estatísticos descritivos. Os resultados desses testes são apresentados na tabela a seguir.

Tabela 1 – Média, desvio padrão e intervalos de confiança das variáveis *monossílabo*, *dissílabo* e *trissílabo* do exercício 1

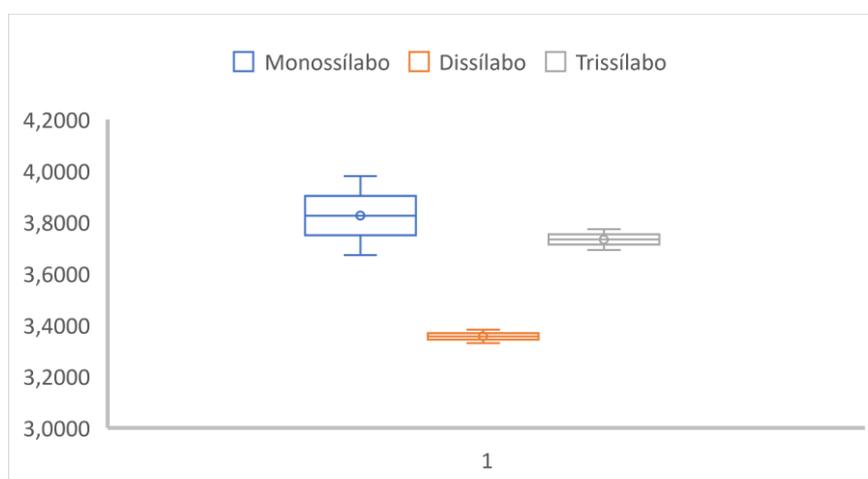
Variáveis	Média	Desvio Padrão	Intervalos de confiança	
			Mínimo	Máximo
Monossílabo	3,83	0,50	3,67	3,98
Dissílabo	3,36	0,08	3,33	3,38

Trissílabo	3,73	0,13	3,69	3,77
------------	------	------	------	------

Fonte: autor (2023)

Entre as médias das variáveis, a que apresenta maior diferença em relação às demais é a variável *dissílabo*. O gráfico de distribuição da variável *monossílabo* já mostrava uma diferença entre os dados e isso se confirma no desvio padrão (0,50), que é o maior entre as variáveis. O gráfico abaixo compara os intervalos de confiança das variáveis. Como já havíamos observado nas médias, os intervalos de confiança apontam para a variável *dissílabo* como a que apresenta maior diferença em relação às outras variáveis. Os intervalos de confiança das variáveis *monossílabo* e *trissílabo* se sobrepõem no gráfico.

Gráfico 2 – Intervalo de confiança das variáveis *monossílabo*, *dissílabo* e *trissílabo* do exercício 1



Fonte: autor (2023)

Para confirmar se a diferença que vemos nos gráficos é significativa, realizamos o teste Friedman, no qual as três variáveis e suas diferenças são testadas. De acordo com o resultado do teste, parece improvável que a diferença encontrada nas respostas dos participantes para os dados de truncamentos de diferentes tamanhos (*monossílabo*, *dissílabo* e *trissílabo*) esteja ocorrendo por erro amostral. ($\chi^2 = 57,98$; $p < 0,001$).

Apesar do resultado do teste Friedman apontar para uma diferença significativa entre as variáveis, verificamos que a variável *monossílabo* possui um número expressivamente menor de dados do que as demais variáveis, pois, no experimento, há apenas duas perguntas com estes dados. Por esse motivo, decidimos realizar novos testes excluindo essa variável para verificarmos se os resultados se alterariam. Realizamos, portanto, um teste Wilcoxon

(equivalente ao teste-t pareado) para verificar a diferença entre as condições *dissílabo* e *trissílabo*.

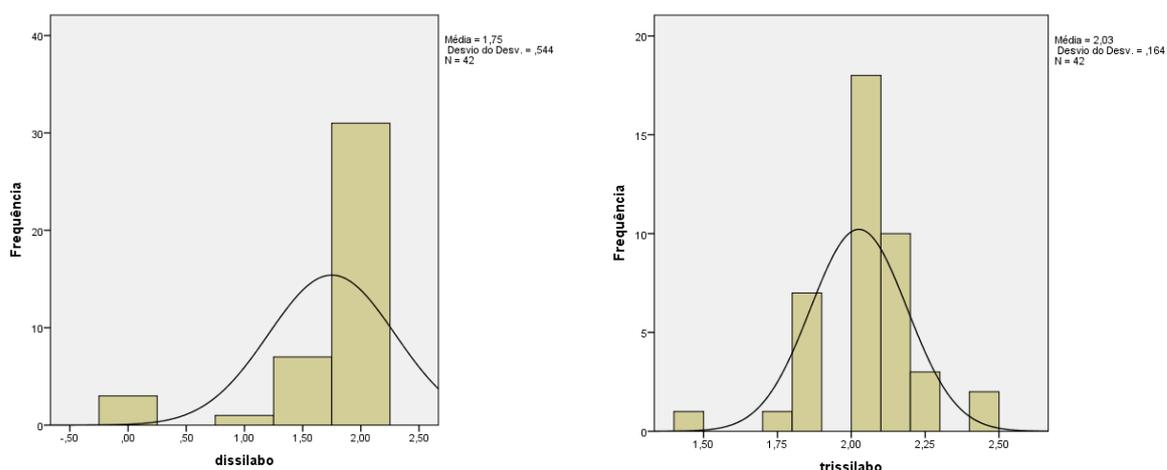
O resultado do teste aponta que há um maior número de sílabas das palavras-base quando há maior número de sílabas na forma truncada. Isso pode ser observado pois o posto médio dos truncamentos dissílabos (0,00) aumentou nos truncamentos trissílabos (22,00). Um escore z de 5,18 tem menos de 5% de probabilidade de ter surgido devido ao erro amostral, caso a hipótese nula seja verdadeira. Por isso, é possível concluir que há um aumento significativo no número de sílabas de palavras-base quando o truncamento é trissílabo em relação a truncamentos dissílabos.

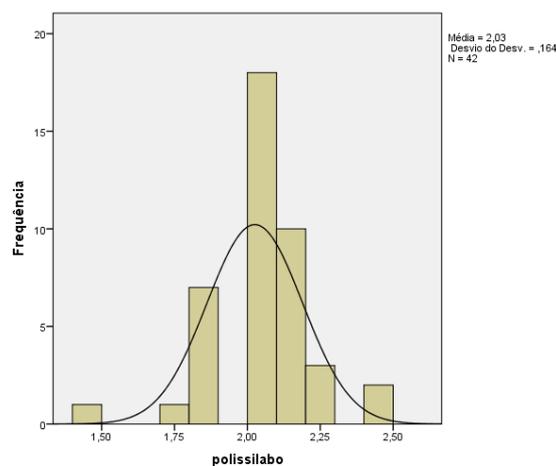
3.2.3.2 Exercício 2

No segundo exercício do experimento, os dados apresentados aos participantes eram de palavras do português e os participantes deveriam propor uma forma truncada para cada uma dessas palavras. Assim como no exercício anterior, dentre as palavras do experimento, havia três variáveis: elas eram dissílabas, trissílabas ou polissílabas e todos os informantes realizavam o experimento nas três condições, por isso temos também aqui um *design* intraparticipantes.

Novamente, verificamos a distribuição dos dados para poder realizar os testes adequados a essa análise. Os gráficos abaixo mostram a distribuição das variáveis.

Gráfico 3 – Distribuição dos dados das variáveis *dissílabo*, *trissílabo* e *polissílabo* do exercício 2





Fonte: autor (2023)

O gráfico da distribuição da variável *dissílabo* aponta para uma concentração maior de dados em palavras com 2 sílabas. A distribuição dos dados nessa variável é assimétrica positivamente. Os gráficos das variáveis *trissílabo* e *polissílabo* apontam para o que parece ser uma distribuição idêntica dos dados nas duas variáveis. Essa distribuição se aproxima um pouco mais da curva de normalidade.

Para obter respostas mais claras quanto à distribuição dos dados, executamos, também, um teste Shapiro-Wilk de normalidade e os resultados apontaram para a distribuição não normal das variáveis (todas as variáveis obtiveram $p < 0,001$). Portanto, se faz necessário o uso de um teste não-paramétrico para a análise dos dados.

Assim como no exercício 1, realizamos testes estatísticos descritivos antes de testes inferenciais. Os resultados estatísticos descritivos dos nossos dados são apresentados na tabela abaixo.

Tabela 2 – Média, desvio padrão e intervalos de confiança das variáveis *dissílabo*, *trissílabo* e *polissílabo* do exercício 2

Variáveis	Média	Desvio Padrão	Intervalos de confiança	
			Mínimo	Máximo
Dissílabo	1,75	0,54	1,58	1,92
Trissílabo	2,03	0,16	1,97	2,08
Polissílabo	2,03	0,16	1,97	2,08

Fonte: autor (2023)

Confirmamos, pela tabela acima, o que já havíamos visto nos gráficos: os resultados das variáveis *trissílabo* e *polissílabo* são idênticas, não há, portanto, diferença entre as mesmas. Por esse motivo, não parece válido realizar um teste Friedman com estes dados. Verificaremos

apenas, se a diferença entre a variável *dissílabo* e as demais é significativa. Para tanto, realizamos um teste Wilcoxon.

O resultado do teste Wilcoxon mostra um aumento no número de sílabas dos truncamentos produzidos pelos participantes quando são apresentados a palavras-base trissílabas e polissílabas. O posto médio das palavras-base dissílaba (8,38) aumentou nas palavras trissílabas e polissílabas (16,06). O escore z foi de 4,01 e o valor de $p < 0,001$. Portanto, é possível concluir que quanto maior a palavra-base, maior a palavra truncada respectiva.

Levando em consideração os resultados dos testes realizados com os dados dos dois exercícios do experimento, a nossa hipótese inicial de que o número de sílabas pode ter papel na formação de truncamentos parece ser confirmada. No exercício 1, encontramos um aumento no número de sílabas da palavra-base quando sua forma truncada era maior. Os resultados do exercício 2 foram semelhantes: quanto maior a palavra-base, maior o truncamento. Esse resultado também vai ao encontro do que Martini (2010) encontrou em seu trabalho, no qual conclui que “formas derivantes mais curtas tendem a truncamentos dissilábicos, ao passo que formas derivantes mais extensas tendem a truncamentos trissilábicos.” (p.243)

É preciso, no entanto, ter um pouco de cautela com os resultados dos testes e conclusões deles advindas. Este estudo não possui uma quantidade muito grande de dados e, provavelmente, por isso, não temos uma distribuição normal dos dados nas variáveis. Como o foco do estudo feito na dissertação não era o de verificar o papel do número de sílabas no truncamento especificamente, não houve uma grande preocupação em equilibrar os dados dentro das variáveis. Por esse motivo, seria necessário realizar novas análises com um maior número de dados e participantes para poder confirmar os resultados aqui encontrados.

3.2.4 Conclusões gerais do estudo sobre o papel do número de sílabas no truncamento

Dentre os estudos sobre truncamento levantados, há poucos que olham para o papel do número de sílabas na formação de truncamentos. Araújo (2002) define que a palavra-base deve ter três sílabas ou mais e que os truncamentos podem ser dissílabos (em sua maioria) e trissílabos. Martini (2010) observou uma tendência em seus dados de que palavras-base mais curtas geralmente formavam truncamentos dissílabos e palavras-base maiores formavam truncamentos trissílabos. O resultado de Martini foi parecido ao encontrado em Heineck (2018) e, a partir disso, formulamos a hipótese de que o tamanho da palavra-base poderia influenciar na extensão de sua forma truncada.

Os resultados das análises realizadas neste trabalho parecem confirmar a nossa hipótese inicial. No exercício 1, verificamos que quando havia um aumento no número de sílabas do truncamento que era apresentado aos participantes, a palavra-base sugerida por eles também aumentava sua extensão. O exercício 2 apresentou um resultado bastante semelhante: nas palavras-base de maior extensão (trissílabo e polissílabo), as formas truncadas produzidas pelos informantes tendem a ser maiores.

Para que possamos ter maior certeza e precisão dos resultados aqui encontrados, precisamos realizar novas análises com um maior número de dados. Os dados utilizados neste estudo não apresentaram uma distribuição normal, muito provavelmente em função da pouca quantidade de dados. As perspectivas de continuidade desse estudo eram de analisar os dados reunidos no *corpus* construído e organizado ao longo do trabalho desta Tese, o que será apresentado na seção a seguir. O objetivo é o de aprimorar a análise e obter resultados mais precisos acerca dos problemas suscitados pelo fenômeno.

3.2 REORGANIZAÇÃO DO *CORPUS*

Para a pesquisa realizada nesta etapa da Tese, os dados do *corpus* foram reorganizados de modo a registrá-los da forma mais completa possível. As anotações fonológicas e morfológicas utilizadas por HEINECK (2018) foram levadas em consideração, mas novas anotações foram incluídas, como, por exemplo, frequência e contexto de uso.

É necessário ressaltar que o levantamento de dados de formas truncadas não é um trabalho fácil e simples. O status informal e marginal de muitas dessas palavras faz com que elas não constem em dicionários e gramáticas. Por isso, a coleta destes dados conta com diferentes ferramentas e é um trabalho que está sempre em progresso. Na construção deste *corpus*, como mencionado anteriormente, o levantamento de dados se deu a partir de estudos anteriores sobre o truncamento (ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014; SCHER, 2018), dos resultados obtidos nos experimentos realizados por Heineck (2018) e do registro de dados encontrados em redes sociais, programas de televisão, etc.

Em Heineck (2018), os dados haviam sido organizados em um arquivo *Word*. Nesta nova etapa da pesquisa, o *corpus* foi transferido para uma tabela *Excel*, de modo a tornar mais

simples a análise dos dados por meio de programas de análise estatística. Até o momento, o *corpus* conta com 239 dados.⁶ Os dados foram decodificados a partir das seguintes anotações:

- i. truncamento
- ii. palavra-base
- iii. número de sílabas do truncamento
- iv. número de sílabas da palavra-base
- v. acento do truncamento
- vi. acento da palavra-base
- vii. frequência em *tokens* do truncamento
- viii. frequência em *tokens* da palavra-base
- ix. composto formado por duas palavras-morfológicas
- x. prefixo autoacentuado
- xi. acréscimo de -a
- xii. acréscimo de -s
- xiii. origem do dado
- xiv. exemplo de uso

No que diz respeito ao registro do truncamento e de sua respectiva palavra-base, é importante destacar que há casos de palavras truncadas que podem ser relacionadas a mais de uma palavra-base (13).

(13) *condi* – condição / condicionador

juve – Juventude / Juventus / juvenil

pinda – pindaíba / Pindamonhangaba

bi – bicampeão / bissexual

Há, ainda, palavras-base com mais de um truncamento possível, como vemos em 14.

(14) *futebol* – fute / futiba / futibas / futebas

⁶ A lista com todos os dados de truncamento está no anexo 10.7.

juvenil – juve / juvena

travesti – trava / traveco

cerveja – ceva / cerva

cafajeste – cafa / cafaja

Esses dados são registrados como entradas diferentes no *corpus*. A frequência, nesses casos, tem um papel importante para identificarmos, no caso de truncamentos com mais de uma palavra-base, qual palavra é mais frequente e, talvez, a que seja mais associada àquela forma truncada pelos falantes. Já no caso de palavras-base com mais de uma forma truncada, é possível verificar qual truncamento tem uma maior frequência de uso.

3.4.1 Número de sílabas

Os dados foram codificados de acordo com o número de sílabas do truncamento e de sua forma truncada. O objetivo é poder observar se o tamanho da palavra-base exerce alguma influência sob o tamanho da sua forma truncada. A classificação foi dividida em *monossílaba*, *dissílaba*, *trissílaba* e *polissílaba*, esta última categoria incluindo palavras com 4 sílabas ou mais.

Na tabela abaixo, é possível observar a distribuição dos dados em relação ao número de sílabas das palavras-base e de suas formas truncadas.

Tabela 3 – Proporção do número de sílabas da palavra-base em relação ao número de sílabas do truncamento

Palavra-base	Truncamento				TOTAL
	Monossílaba	Dissílaba	Trissílaba	Polissílaba	
Dissílaba	0 0%	10 4,18%	0 0%	0 0%	10 4,18%
Trissílaba	0 0%	52 21,76%	16 6,69%	0 0%	68 28,45%
Polissílaba	7 2,93%	87 36,40%	65 27,20%	2 0,84%	161 67,36%
TOTAL	7 2,93%	149 62,34%	81 33,89%	2 0,84%	239 100%

Fonte: autora (2023)

Ao observarmos a distribuição dos dados de truncamentos, as formas truncadas dissílabas são as mais comuns no *corpus* e compõem 62,34% do total de dados. Logo em

seguida, estão os truncamentos trissílabos, que totalizam 33,89% dos dados. Os únicos sete dados (2,93%) de truncamentos monossílabos têm como origem palavras-base compostas por duas palavras fonológicas com prefixos autoacentuados, como *pré-escola*, *pós-graduação* e *ex-namorado*. Há apenas dois dados (0,84%) de truncamentos polissílabos, um deles é formado por um composto neoclássico — *otorrinolaringologista* > *otorrino* — e o outro tem como base uma palavra composta, mas seu truncamento é formado de parte das duas palavras do composto — *Porto Alegre* > *Portoale*.

Apenas 10 dados (4,18%) do *corpus* são de palavras-base dissílabas e todos formam truncamentos também dissílabos. As palavras-base trissílabas compõem 28,45% por cento dos dados do *corpus*, com 68 dados e grande parte dos truncamentos formados a partir destas palavras são dissílabos — 52 dados, correspondendo a 21,76% do *corpus*. Temos ainda exemplos de palavras trissílabas que são base para a formação de truncamentos trissílabos, mesmo que em menor quantidade (16 dados – 6,69%).

Quando a palavra-base tem quatro sílabas ou mais, os truncamentos são, em sua grande maioria, dissílabos e trissílabos. Assim como quando as palavras-base são dissílabas e trissílabas, aqui também observamos que os truncamentos dissílabos predominam (87 dados – 36,40%). No entanto, aqui temos uma quantidade considerável de truncamentos trissílabos: 81 65 dados (27,20%).

A distribuição dos dados em relação ao número de sílabas dos truncamentos e suas formas truncadas observada no *corpus* se assemelha ao que foi encontrado por Heineck (2018) nos resultados dos experimentos (cf. seção 3.1). Assim como nos resultados da dissertação, no *corpus*, há uma tendência de o truncamento ser menor que a palavra-base, visto que isso ocorre em 85,36% dos dados do *corpus*. Os dados do *corpus*, assim como os resultados da dissertação, também corroboraram a tendência proposta por Martini (2010) que aponta que, geralmente, palavras-base dissílabas formam truncamentos dissílabos, palavras-base trissílabas formam truncamentos dissílabos e, por fim, palavras-base polissílabas formam truncamentos trissílabos.

Os resultados dos experimentos de Heineck (2018) foram submetidos a uma análise estatística inferencial para verificar se o número de sílabas da palavra-base poderia ter alguma influência no tamanho de sua respectiva forma truncada (cf. subcapítulo 3.2). A análise dos dados confirmou a hipótese inicial e apontou para uma correlação entre o tamanho da palavra-base e o tamanho do truncamento, ou seja, quanto maior a palavra-base, maior é sua forma truncada.

Observando a distribuição dos dados do *corpus*, essa correlação não parece se confirmar, visto que a maioria dos truncamentos é dissílaba, independentemente do número de sílabas da palavra-base. Percebemos, no entanto, que há um aumento dos dados de truncamentos trissílabos quando a palavra-base é polissílaba, mas isso pode ser devido à tendência de a forma truncada ser menor que a sua base e, por esse motivo, haveria mais dados de truncamentos trissílabos com palavras polissílabas do que com palavras-base trissílabas.

3.4.2 Acento

Os dados do *corpus* também foram codificados de acordo com a posição do acento e, portanto, foram classificados em *oxítono*, *paroxítono* e *proparoxítono*. Na tabela 4, é possível ver a distribuição dos dados de acento da palavra-base em relação ao acento do truncamento.

Tabela 4 – Proporção do acento da palavra-base em relação ao acento do truncamento

Palavra-base	Truncamento		
	Oxítono	Paroxítono	TOTAL
Oxítona	18 7,53%	55 23,01%	73 30,54%
Paroxítona	36 15,06%	127 53,14%	163 68,20%
Proparoxítona	1 0,42%	2 0,84%	3 1,26%
TOTAL	55 23,01%	184 76,99%	239 100%

Fonte: autora (2023)

Nos dados do *corpus*, há apenas truncamentos oxítonos e paroxítonos; não temos registro de truncamentos proparoxítonos. Mesmo entre as palavras-base, há apenas três dados de palavras proparoxítonas. Grande parte das formas truncadas é paroxítona, compondo 76,66% do total de dados. Essa distribuição dos dados vai ao encontro do que foi encontrado em Heineck (2018), em que a maioria dos dados era dissílaba e paroxítona. A formação de truncamentos paroxítonos é esperada, visto que esse é o padrão acentual mais regular da língua portuguesa. Não parece haver, portanto, uma relação entre o padrão acentual da palavra-base e de seu truncamento.

3.4.3 Frequência

A frequência dos dados foi atestada através da consulta ao Corpus do Português, no subcorpus *Web/Dialects*⁷. Este subcorpus contém cerca de um bilhão de palavras que foram coletadas em páginas da internet de 4 países falantes de português (Brasil, Portugal, Angola e Moçambique) entre os anos de 2014 e 2022.

Verificamos a frequência tanto das formas truncadas quanto de suas palavras-base e restringimos a busca apenas a dados do Brasil. A vantagem de utilizar os dados do subcorpus *Web/Dialects* está no fato de sua base de dados vir de sites de notícias, de esportes, moda, blogs pessoais, etc. e incluir dados de comentários feitos pelos leitores destes sites. Para uma pesquisa que lida com dados informais e pouco frequentes como são os truncamentos, esse tipo de *corpus* pode ser mais interessante do que o de corpora que lidem apenas com textos escritos mais formais.

A análise dos resultados obtidos na verificação de frequência dos dados será realizada na parte 2 deste trabalho, no capítulo 5.

3.4.4 Palavra-base formada por duas palavras morfológicas e/ou duas palavras fonológicas

Nesta seção, tratamos de duas classificações às quais os dados do *corpus* foram submetidos: palavras-base formadas por duas palavras morfológicas — diz respeito a palavras compostas por duas raízes eruditas, como *fotografia* — e palavras-base formadas por duas palavras fonológicas — engloba palavras compostas por prefixos autoacentuados, como *pós-graduação*.⁸

Dos 239 dados que compõem o *corpus*, 38 (15,90%) são de truncamentos com palavras-base compostas por duas raízes eruditas, como em *odontologia* > *odonto*, *psicólogo* > *psico*. O entendimento deste tipo de dado como resultado do processo de truncamento é bastante discutido na literatura. Alguns autores excluem esses dados da análise por compreenderem que eles resultam de processos de formação de palavras diferentes do truncamento.

Araújo (2002) classifica dados como *eletrocardiograma* > *eletro* como pseudotruncamentos, pois, de acordo com o autor, eles são derivados de afixos cultos de origem greco-latina, germânica, etc., e a forma encurtada — *eletro* — é identificada e usada metonimicamente. Vilela, Godoy e Silva (2006) concordam, em parte, com Araújo (2002) e

⁷ Disponível em <https://www.corpusdoportugues.org/web-dial/>

⁸ É importante destacar que o limite morfológico, que delimita o prefixo e as raízes eruditas, coincide, em muitas das vezes, com o limite da palavra fonológica.

diferenciam o processo do truncamento do processo de redução vocabular. As autoras afirmam que a redução vocabular seria um fenômeno apenas morfológico, pois o local de corte na palavra que sofre a redução é o limite morfológico, como em *fotografia* > *foto* e, portanto, este não seria um processo morfofonológico como o truncamento.

Martini (2010) também não considera na sua análise formas que se estruturam com a manutenção integral do morfema à esquerda da forma de base, como ocorre com os compostos formados por raízes eruditas. A autora, de forma semelhante ao trabalho de Vilela, Godoy e Silva (2006), define este processo como puramente morfológico.

De forma diferente aos autores já citados, Scher (2011 e 2018) inclui este tipo de dado a sua análise, classificando-o como um dos padrões de formação de truncamentos – Padrão II (SCHER, 2011) e Tipo A (SCHER, 2018). Belchor (2014) também classifica esses dados como um dos três padrões de formação de truncamentos: o Padrão Odonto. Ambas as autoras também incluem nesses padrões os dados de truncamentos formados a partir de prefixos autoacentuados (ex.: pós, ex, vice).

Na composição do *corpus*, decidimos incluir estes dados, considerando, assim como Scher e Belchor, que eles podem ser apenas um tipo ou padrão de formação de truncamentos. Mesmo assim, acreditamos que é importante classificá-los, de modo que, em uma análise, seja possível observar se eles se comportam de maneira diferente aos demais dados, como observou Schwindt (2000), que propôs que esses prefixos são palavras fonológicas independentes e, por isso, têm autonomia prosódica.

Além das palavras compostas, também classificamos os dados derivados de palavras compostas por sufixos autoacentuados. Esses dados compõem apenas 5,05% do *corpus* (12 dados). Nesta categoria, é importante destacar o caso de três formas truncadas: *ex*, *bi* e *vice*, pois eles podem ter como base mais de uma palavra, como vemos em (15).

(15) *ex* – *ex-namorado* / *ex-marido*

bi – *bissexual* / *bicampeão*

vice – *vice-presidente* / *vice-governador* / *vice-prefeito* / *vice-campeão*

Na busca por estas formas truncadas no Corpus do Português, verificamos uma alta frequência de tokens e, por esse motivo, a identificação de qual a palavra-base de cada dado se tornou muito difícil. Desse modo, registramos apenas a frequência de *ex*, por exemplo, sem separar quantos tokens eram relacionados a *ex-namorado* e quantos se referiam a *ex-marido*. A frequência de cada uma das palavras-base, no entanto, foi verificada separadamente.

3.4.5 Acréscimo de -a

O acréscimo de uma vogal -a em algumas formas truncadas é um fato atestado e bastante discutido nos diferentes estudos sobre o truncamento. Por esse motivo, acreditamos na importância de incluir essa anotação aos dados do nosso *corpus*. Dos 239 truncamentos que compõem o *corpus*, atestamos o acréscimo de -a em 69 dados, ou seja, 28,87% do total. Na sequência, iremos retomar as diversas discussões já feitas sobre essa vogal acrescida às formas truncadas.

Araújo (2002) trata do acréscimo da vogal -a apenas aos truncamentos trissílabos. O autor afirma que essa vogal é uma espécie de *epêntese* que tem por função impedir que se produzam formas agramaticais (*portug para ‘português’) e, ao mesmo tempo, impede que se apague mais material fonológico da palavra-matriz, o que poderia provocar a perda de lexicalidade da palavra truncada.

Gonçalves (2004) trata de casos nos quais as formas truncadas são resultado da reprodução de parte da base e do acréscimo de uma vogal -a que nem sempre está presente na palavra-matriz, como em *vestiba*, por *vestibular*, *estranja*, por *estrangeiro*, e *sarja*, por *sargento*. O autor afirma que a vogal -a funciona como “uma espécie de afixo de Truncamento, que, por isso, pode ser considerado processo simultaneamente não concatenativo (cópia) e aglutinativo (acréscimo de vogal final)” (GONÇALVES, 2004, p.13).

Vilela, Godoy e Silva (2006), em um artigo no qual revisam os trabalhos de Gonçalves (1999) e Araújo (2002), reservam uma seção do artigo que chamam de “vogal temática”, que trata justamente da vogal -a que é acrescida às formas truncadas. No entanto, as autoras não dão maiores explicações das razões da escolha dessa nomenclatura para classificar essa vogal. A seção é destinada a questionamentos à afirmação de Gonçalves (1999) de que há sempre a inserção de uma vogal -a no truncamento. As autoras afirmam que, para os truncamentos trissilábicos, a generalização da vogal final -a pode valer como tendência, apesar de haver alguns contraexemplos. Já para os casos dos truncamentos dissilábicos, a vogal -a também é acrescida em alguns casos, mas há, porém, nos dados das autoras cerca de 50% de contraexemplos à generalização proposta por Gonçalves (2009). Apesar de não apresentarem argumentos para a escolha de classificar a vogal final -a como *vogal temática*, as autoras podem tê-la assim classificado, pensando que essa vogal -a não teria a função de marcar gênero. Em um truncamento como *vestiba* (para *vestibular*), por exemplo, a forma truncada é masculina — o *vestiba* —, e do mesmo modo que com uma palavra como *o planeta*, o -a deve ser entendido como vogal temática.

Santos (2002) destaca o *status* morfopragmático da vogal final -a, pois ela pode expressar pejoratividade, como em *japa*, *trava*, ou afetividade, como em *Sampa*, *Floripa*, *cerva*. Essa função

expressiva pode também ser atribuída a uma vogal final -o, em formas truncadas como *milico* < *militar*, *boteco* < *botequim*, *motô* < *motorista*, *traveco* < *travesti*. Diferentemente das vogais das formas truncadas, as vogais temáticas dos fenômenos concatenativos não têm essa propriedade expressiva dos truncamentos e, por esse motivo, que elas seriam mais próximas de sufixos derivacionais.

Para tratar das formas truncadas do Tipo III, Scher (2016) baseia-se na proposta de Alcântara (2003, 2010) de morfemas de classe. De acordo com Alcântara, substantivos e adjetivos são divididos em três classes: classe I, constituída de nomes terminados em -o; classe II, constituída de nomes terminados em -a; e a classe III, formada por nome terminados em -e ou outro segmento. Scher observa que, diferentemente de suas palavras-base, que são nomes com diferentes terminações e, portanto, pertencentes a diferentes classes, a maioria dos truncamentos em PB são terminados em -a e, portanto, parecem pertencer à classe II.

3.4.6 Acréscimo de -s

Estudos mais recentes têm lançado um olhar para casos de truncamento em que há o acréscimo da consoante -s em algumas formas truncadas. Em nosso *corpus*, foram registrados 21 dados (8,79%) em que o -s foi acrescido ao truncamento, como em *futebol* > *futibas*, *tranquilo* > *tranquis* e *bobeira* > *bob(i)s*.

Scher (2011) inclui dados de truncamento em que há o acréscimo de -s em um padrão em que a forma truncada corresponde a uma palavra bimorfêmica, ou interpretada como tal pelo falante: raiz, ou parte da raiz da palavra correspondente + sequência (V)C $-(a)s$ ou $-(i)s$ ou (V)CV $-(u)ca$ (*migs* < *amigo*; *mamis* < *mamãe/mãe*; *legas* < *legal*; *feijoca* < *feijoada*).

Scher (2016) retoma seu estudo anterior e levanta uma discussão sobre as formas -as e -is a partir das classes morfêmicas formais propostas por Alcântara (2010). Como primeira hipótese, a autora propõe que essas formas façam parte da Classe II, que abrange as palavras terminadas em a, tanto femininas quanto masculinas, e que, portanto, -as e -is seriam alomorfes de a. De acordo com a autora, essa hipótese se sustenta no fato de que, apesar de ser foneticamente nulo, [EVAL] contém traços idiossincráticos da classe II, que vão exigir a inserção de itens de vocabulário dessa classe no morfema abstrato, que irá corresponder ao sufixo temático na estrutura do truncamento (\mathfrak{S}). Seria natural, portanto, com a inserção de [EVAL], que tanto o a quanto -as e -is fossem inseridos na palavra.

Apesar de essa ser uma hipótese que parece bastante consistente, Scher ainda considera uma discussão levantada por Alcântara (2003) na sua análise. Nesse trabalho, Alcântara propõe

uma nova classe — Classe V —, que engloba palavras como *oásis*, *vírus* e *brócolis*. Essas palavras têm a mesma forma para o singular e o plural e não necessitam do acréscimo de *e* para a formação do plural, como ocorre em *capuz-capuzes*. Os truncamentos terminados em *-as* e *-is* têm essa mesma característica, ou seja, não é possível formar o plural a partir do acréscimo de *-e* (*bermas* - **bermases*). Outra característica das formas da Classe V é o fato de que o segmento *-s*, que aparece no final de *vírus* e *oásis*, não fazer parte da palavra, pois na derivação ele desaparece (*vírus* – *viral* - **virsal*). O mesmo acontece com os truncamentos em que *-as* e *-is* que também não fazem parte da raiz da palavra (*bermas* – *bermuda* - **bermasmuda*).

A questão levantada por Scher é que, ao considerar *-as* e *-is* como pertencentes à Classe V, teríamos que explicar por que, na presença de [EVAL], a vogal *a* da Classe II não é inserida, como é exigido por [EVAL]. E se, por outro lado, *-as* e *-is* pertencem à Classe II (e são alomorfes de *a*), de que forma eles se distinguem das palavras da Classe V?

3.4.7 Origem do dado e exemplo de uso

Cada entrada do *corpus* tem o registro da origem do dado. As anotações indicam se o dado foi retirado de um estudo sobre o processo do truncamento, com a devida referência, ou se foi encontrado pela autora em redes sociais, programas de televisão ou produzidos em situações de interações de falantes.

Além do registro da origem do dado, também incluímos um exemplo de uso do item. Os exemplos foram retirados do Corpus do Português e incluem o contexto de uso do dado, que é importante para que possamos confirmar seu significado, sua palavra de origem e para que possamos analisar os contextos em que o dado pode ser utilizado. Para itens com zero ocorrências no Corpus do Português, o exemplo de uso pode ter como origem sites da internet, redes sociais, aplicativos de mensagem instantânea ou, se não havia um contexto real de uso, foram criados pela autora.

PARTE 2

PAPEL DA FREQUÊNCIA LEXICAL NO TRUNCAMENTO

5 FREQUÊNCIA LEXICAL

5.1 FONOLOGIA DE USO

Bybee (2004, p.1) 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 palavras. A autora 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. Nesta proposta, 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 seguintes princípios básicos.

1. Experiência afeta 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 têm 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 de que elas 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, p. 6, 7 e 8)

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 (2013) 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” (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.

6 ESTUDO DE FREQUÊNCIA COM DADOS DO *TWITTER* E DO *CORPUS* DO PORTUGUÊS

6.1 HEINECK (2021)

Dando sequência ao trabalho realizado em minha dissertação de mestrado (HEINECK, 2018), realizamos um estudo com o objetivo de verificar a frequência de uso de dados de truncamento obtidos nos experimentos realizados no estudo anterior. A verificação foi realizada na rede social *Twitter*, a partir de um pacote do programa computacional R — o pacote *rtweet*.

Este trabalho (HEINECK, 2021) teve como principal objetivo investigar se a frequência tem algum papel no processo de truncamento. Ao verificar a frequência de ocorrência de formas truncadas, objetivou-se 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*. Nas próximas seções, detalharemos de que forma foi realizada a seleção dos dados, a busca destes dados no *Twitter*, bem como os principais resultados obtidos.

6.1.1 Seleção dos dados

Os dados de truncamento analisados neste estudo foram dados retirados dos resultados dos dois experimentos (*online* e presencial) realizados em Heineck (2018). 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 apontaram *japonês* como palavra-base; por isso, esse dado compôs a lista das formas truncadas que foram buscadas no *Twitter*. O objetivo era verificar se esses truncamentos em que a base é facilmente recuperável pelos falantes são bastante frequentes na língua e, dessa forma, formular hipóteses

sobre uma possível lexicalização destes truncamentos: 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, era verificar se os truncamentos mais frequentes nos resultados são também frequentes no *Twitter*. Decidimos diminuir a porcentagem de ocorrência em relação aos dados do exercício 1, pois, no exercício 2, houve poucos dados com frequência maior do que 70%, devido a característica do exercício, em que os participantes poderiam criar formas truncadas.

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. Esses dados contribuem para a discussão acerca do *status* da vogal final do truncamento, pois são truncamentos femininos terminados em *o* que derivam de palavras terminadas em *a* — *fotografia* e *motocicleta*. A escolha destes dados se deu a partir dos resultados de Heineck (2018), em que as formas diminutivas de *foto* e *moto* escolhidas pelos participantes dos experimentos, foram as formas terminadas em *o* — *fotinho* e *motinho*. Pretendemos verificar se o mesmo ocorre nos dados do *Twitter*.

6.1.2 Busca no *Twitter*

Após a seleção dos dados, realizou-se uma busca no *Twitter* dos dados selecionados. Para tanto, utilizou-se um pacote do programa computacional R — o pacote *rtweet*. Com esta ferramenta, é possível que sejam feitas pesquisas no *Twitter* a partir de alguma palavra, frase ou a partir de algum tópico que esteja em alta no momento (os chamados *trending topics*). 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 na versão que utilizamos. 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 aumentando a frequência de algum dado. Realizaram-se três buscas com um intervalo de tempo de, no mínimo, uma semana entre elas,

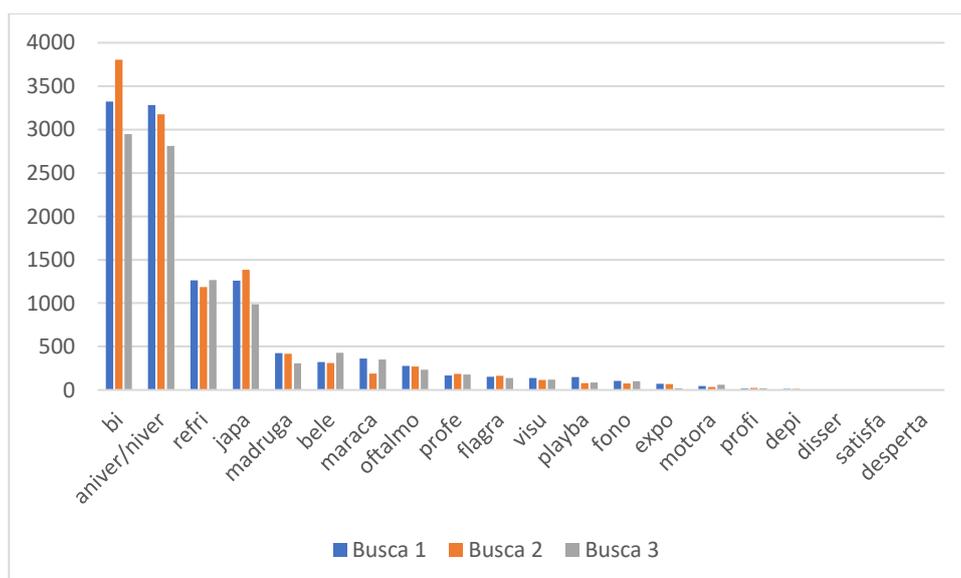
de modo a verificar se não haveria uma discrepância na frequência de ocorrência algum dado nas buscas.

6.1.3 Principais resultados

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 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 4 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 4 – Frequência de ocorrência dos dados do exercício 1 nas três buscas realizadas no *Twitter*



Fonte: Heineck (2021)

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.

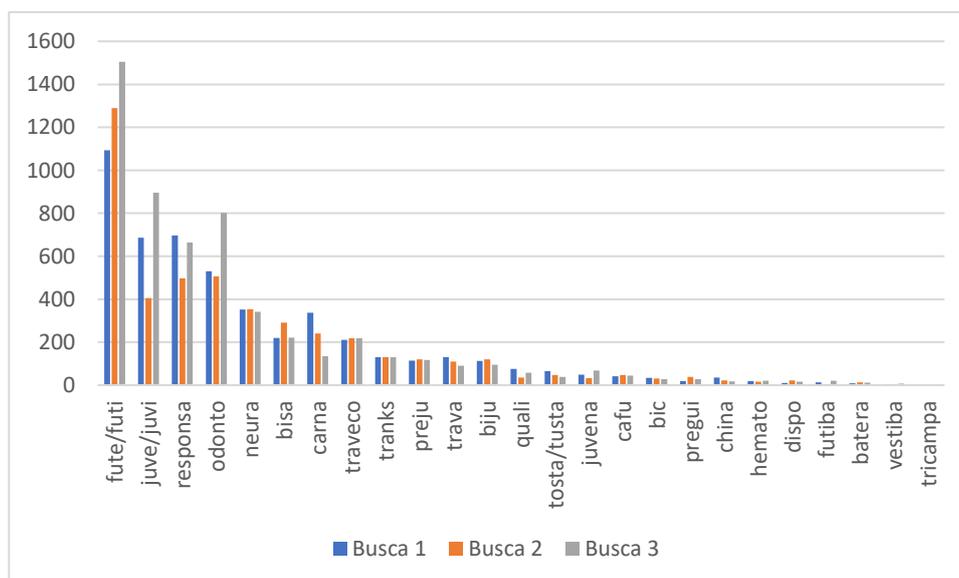
Tabela 5 – Dados do exercício 1 com variação de frequência de ocorrência maior que 20% em relação à média da frequência nas três buscas realizadas no *Twitter*

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

Fonte: Heineck (2021)

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



Fonte: Heineck (2021)

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.

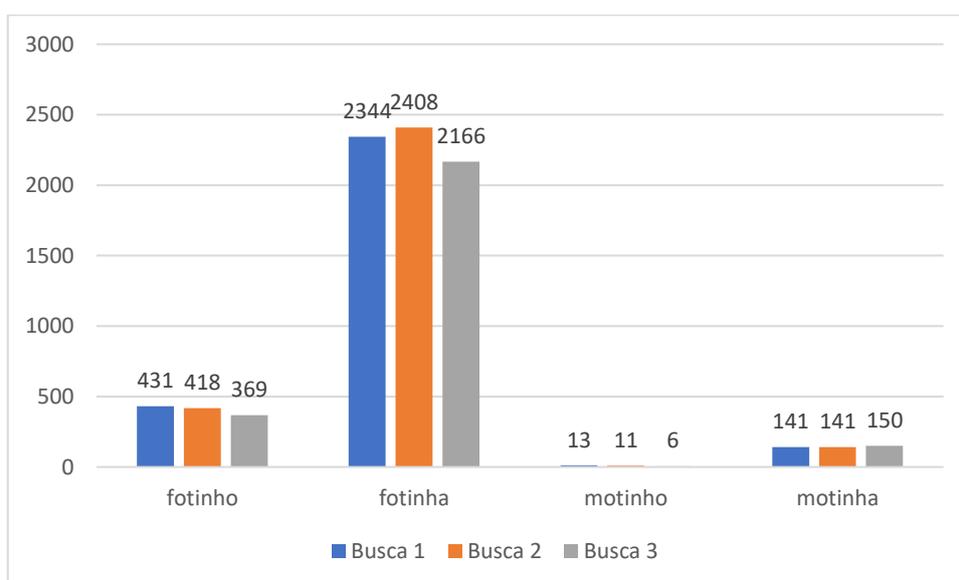
Tabela 6 – Dados do exercício 2 com variação de frequência de ocorrência maior que 30% em relação à média da frequência nas três buscas realizadas no *Twitter*

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

Fonte: Heineck (2021)

Por fim, o Gráfico 6 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 6 – Frequência de ocorrência dos dados do exercício 3 nas três buscas realizadas no *Twitter*



Fonte: Heineck (2021)

Nas subseções a seguir, será realizada uma análise mais cuidadosa dos dados de frequência dos truncamentos de cada um dos exercícios.

6.1.3.1 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 5. Calculamos a média dos resultados das três buscas realizadas para organizarmos os dados em mais e menos frequentes.

Tabela 7 – Dez truncamentos do exercício 1 mais frequentes no *Twitter*

Truncamento	<i>Twitter</i> Frequência (média)	Concordância na escolha da palavra-base	
		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 %
madrug	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 %

Fonte: Heineck (2021)

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*, *madruga*, *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 (SCHWINDT, 2000), *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) que exclui de sua análise formas “que se estruturam com a manutenção integral do morfema à esquerda da forma de base” (p. 254), 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.

6.1.3.2 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 8 – Dez truncamentos do exercício 2 mais frequentes no *Twitter*

Truncamento	<i>Twitter</i> Frequência (média)	Ocorrência da forma truncada	
		Experimento <i>on-line</i>	Experimento presencial
fute/futi	1296	67,4 %	87 %
juve/juvi	662,3	62,8 %	37 %
respon	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 %

Fonte: Heineck (2021)

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.

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

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

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

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

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

Nos exemplos (16) e (17), *tosta* tem o sentido de *torrada* ou *misto-quente*, em (17), inclusive, *tosta* está acompanhado de *mista*, isso aparece com bastante frequência nos dados. Já nos exemplos (18) e (19), *tosta* tem o sentido de *calor* ou *sol muito forte/quente*. Por fim, no exemplo (20), *tosta* está caracterizando *novela* e parece ter o sentido de “ruim”, “de mau 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.

6.1.3.3 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 7. *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 9 – Dados de frequência dos diminutivos de *foto* e *moto* no *Twitter*

Truncamento	<i>Twitter</i> Frequência (média)	Ocorrência da forma truncada	
		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%

Fonte: Heineck (2021)

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.

6.1.4 Conclusões gerais do estudo sobre frequência no *Twitter*

Os resultados da busca realizada no *Twitter* apontam que os truncamentos nos quais os falantes recuperavam com facilidade a base nos experimentos realizados em minha dissertação 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 de Heineck (2018) 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.

6.2 FREQUÊNCIA COM DADOS DO *CORPUS* DO PORTUGUÊS

De modo a dar continuidade ao estudo feito em Heineck (2021) com dados do *Twitter*, realizamos a busca dos dados de truncamento e de suas respectivas palavras-base no *Corpus* do Português. Essa busca fez parte da construção do nosso *corpus* de dados de truncamento, que já conta com 239 dados⁹ que foram codificados de acordo com características fonológicas,

⁹ A lista com todos os dados pode ser encontrada no Anexo 10.7.

morfológicas, como sílaba e acento, por exemplo, e que contam com informações sobre sua origem e exemplos de uso.

Figura 7 – Exemplo de codificação dos dados do *corpus*

Truncamento Base	Sílaba_Trunc	Sílaba_Base	Acento_Trunc	Acento_Base	Frequência_Trunc	Frequência_Base	Composto	Prefixo_autc	Acréscimo_a	Origem	Exemplo_Uso
paleo	paleontolc	2	4	paroxítono	paroxítono	4	351 s	n	n	B	Um estudo publicado em a revista científica " Paleo " mostra uma análise indicando que os saurípedes, dinossauros herbívoros de
psico	psicólogo	2	4	paroxítono	proparoxítona	8	9456 s	n	n	B, S	eh algo que mesma programei, inconscientemente. Minha psico falou q preciso me acelar de o jello que eu sou, mas como
penta	pentacamj	2	4	paroxítono	oxítono	301	227 s	n	n	B	O Golária, em os bons tempos, fo penta entre 1991 e 1994. Mas o o título de este texto? Como
hetero	heterossex	3	4	paroxítono	oxítono	1292	1875 s	n	n	B, VGS	com o dia de que se uma pessoa se hetero e faz sexo com pessoas de outro sexo automaticamente ela é hetero. Sei
homo	homossexi	2	4	paroxítono	oxítono	543	9490 s	n	n	B, VGS	lutar por a aprovação em o congresso por o casamento homo e adoção de crianças por tais casais se deve a pressão de o Vaticano
hemato	hematolog	3	4	paroxítono	paroxítono	5	149 s	n	n	B	batalha e sofrimento com meu filho e tratamentos, o hemato parou com corticoides a muitos anos e falou para ele tomar esta vitamina
tetra	tetracamp	2	4	paroxítono	oxítono	192	212 s	n	n	B	a postura escavocada e desleigante de o capião de o Tetra de 1994 e atual letreiro de o Intercontinental, Dunga, que realmente tá
neuro	neurologis	2	4	paroxítono	paroxítono	130	1677 s	n	n	B	de o que pode ser o seu diagnóstico, o neuro não falou nada a respeito de isso, apenas os traços autísticos.
esteto	estetoscóç	3	4	paroxítono	paroxítono	15	340 s	n	n	B	necessários, salvando muitas vidas. Poderá diagnosticar pneumonia com esteto , ou seja, já em fase avançada, dando antibióticos
otorrino	otorrinolai	4	4	paroxítono	paroxítono	431	363 s	n	n	B, VGS	onstante em a orelha direita. Fiz audiometria e o otorrino disse que não tenho nenhuma perda de audição. Não consigo me concentra
lipo	lipoaspiraç	2	4	paroxítono	oxítono	470	512 s	n	n	B, VGS	para replantar em o bumbum. "x Kim fez uma lipo em as pernas e o médico retirou a gordura de as pernas replicando em
super	supermerc	2	4	paroxítono	paroxítono	11	7672 s	n	n	S	frete com para a nutricionista, e endócrino e o super para comprar as coisas de o regime, tenho feito musculação 5 vezes em
curta	curta-metr	2	4	paroxítono	paroxítono	342	853 s	n	n	VGS	O cinema brasileiro é feito com dinheiro público. Com o dinheiro de um curta se fazem quatro casas populares, com o dinheiro dum

Fonte: Heineck (2023)

A busca dos dados foi realizada no subcorpus *Web/Dialects* que conta com dados de 4 países falantes de português (Brasil, Portugal, Angola e Moçambique) coletados de páginas da internet entre os anos de 2014 e 2022. Acreditamos que, assim como o *Twitter*, um *corpus* com dados da internet pode ser um local de pesquisa muito rico para o tipo de dado que buscamos. Os dados de truncamento dificilmente serão encontrados em abundância em textos formais e, portanto, *corpus* com dados escritos que se aproximam mais da fala informal, como é o caso do *Twitter* e do *Web/Dialects* são interessantes.

Ampliamos a busca que havia sido feita no *Twitter* para todos os dados do nosso *corpus* e incluímos a verificação da frequência das palavras-base. Desse modo, é possível construir hipóteses mais robustas acerca do papel da frequência no processo de formação e uso do truncamento.

Nas próximas seções, iremos apresentar os principais resultados da busca no Corpus do Português e estabeleceremos comparações com o estudo realizado com dados do *Twitter*.

6.2.1 Principais resultados

Tendo em vista o foco do nosso estudo, a busca dos dados foi limitada a dados do Brasil. Dentre os dados de truncamentos, a frequência variou de 0 a 112.644, e a média foi de 1.152,42. Assim como havíamos verificado na busca no *Twitter*, a realidade do processo do truncamento faz com que alguns dados sejam difíceis de encontrar em corpora de dados escritos. Em nosso *corpus*, 73 itens de formas truncadas tiveram 0 ocorrências no *Web/Dialects*, a lista pode ser encontrada no Anexo 10.4.

A tabela abaixo apresenta a distribuição da frequência dos dados de truncamento do nosso *corpus*.

Tabela 10: Distribuição da frequência em *tokens* dos dados de truncamento

Ocorrência no <i>Web/Dialects</i> em <i>tokens</i>	Quantidade de itens do <i>corpus</i>
+ de 100.000	1
10.001 a 100.000	4
1.001 a 10.000	11
501 a 1.000	10
101 a 500	33
1 a 100	103
0	73

Fonte: autor (2023)

Grande parte dos nossos dados se concentram na faixa entre 1 e 100. O único dado com ocorrência maior do que 100.000 é *foto* com 112.644 ocorrências. Olhamos mais atentamente para os dados com mais de 10.000 ocorrências e comparamos com a ocorrência de suas respectivas palavras-base.

Tabela 11: Dados de truncamento com mais de 10.000 ocorrências

Truncamento	Ocorrência	Palavra-base	Ocorrência
foto	112644	fotografia	19119
ex	25710	ex-namorado	1743
ex	25710	ex-marido	2340
moto	12030	motocicleta	2590
vice	11473	vice-presidente	11137
vice	11473	vice-governador	2432
vice	11473	vice-campeão	700
vice	11473	vice-prefeito	2699
pós	10393	pós-graduação	10480

Fonte: autor (2023)

Os dados de truncamentos mais frequentes não correspondem às palavras-base mais frequentes no *Web/Dialects*, como veremos a seguir. No entanto, todas as palavra-base apresentaram mais de 1.000 ocorrências, com exceção apenas de *vice-campeão*, com 700 ocorrências. Vale ressaltar ainda os casos de palavra-base com mais de 10.000 ocorrências: *fotografia* (19.119), *vice-presidente* (11.137) e *pós-graduação* (10.480).

A tabela abaixo apresenta a distribuição dos dados de palavra-base de acordo com sua ocorrência no *corpus Web/Dialects*.

Tabela 12: Distribuição da frequência em *tokens* dos dados de palavras-base

Ocorrência no <i>Web/Dialects</i> em <i>tokens</i>	Quantidade de itens do <i>corpus</i>
100.000 a 600.000	13
10.001 a 100.000	63
1.001 a 10.000	94
501 a 1.000	19
101 a 500	40
1 a 100	9

Fonte: autor (2023)

A maior concentração de itens de palavras-base está na ocorrência de 1.001 a 10.000 no *Web/Dialects*. Diferentemente dos dados de truncamentos, apenas poucos dados apresentaram ocorrência entre 1 e 100, os dados em que isso ocorreu estão listados abaixo

estrogonofe – 92 ocorrências

balzaquiana – 82 ocorrências

Morro da Providência - 70 ocorrências

overnight – 59 ocorrências

jabaculê – 45 ocorrências

espeleologia – 36 ocorrências

grã-fino – 30 ocorrências

mocreia – 22 ocorrências

pornográfico – 2 ocorrências

Olhamos para as palavras-base mais frequentes do nosso *corpus* e comparamos com a frequência de suas respectivas formas truncadas.

Tabela 13: Dados de palavras-base com mais de 100.000 ocorrências

Truncamento	Ocorrência	Palavra-base	Ocorrência
Brasil	591.398	brasa	10
trabalho	468.931	trabis	271

mulher	265.158	mulha	0
primeira	260.249	prima	1
presidente	196.451	presida	12
coração	175.140	coras	0
coração	175.140	coraça	2
vontade	160538	vontis	0
internet	156.227	interne	25
justiça	145.734	justa	0
amigo	122.626	miguís	9
presente	116.207	preza	0
natureza	111.714	natura	1
profissional	101.191	profí	10

Fonte: autor (2023)

Grande parte dos dados de truncamento (10 itens) das palavras-base mais frequentes tem entre 0 e 10 ocorrências e, desses, 5 itens tem 0 ocorrências no *corpus Web/Dialects*. O único item com mais de 100 ocorrências é o truncamento *trabis* (271 ocorrências). Anteriormente, havíamos observado que os truncamentos mais frequentes também tinham palavras-base frequentes. No entanto, aqui, percebemos que palavras-base muito frequentes não têm, necessariamente, truncamentos frequentes.

6.2.2 Comparação com dados do *Twitter*

Comparamos os resultados de Heineck (2021) que realizou a busca dos dados no *Twitter* com os resultados obtidos na busca no *Web/Dialects*. A tabela abaixo apresenta a média do número de ocorrência dos 10 truncamentos mais frequentes no *Twitter* e o número de ocorrências do mesmo item no *Web/Dialects*. É importante salientar que os dados utilizados por Heineck (2021) eram resultados dos experimentos realizados por Heineck (2018) e, por isso, estão divididos entre exercício 1 e exercício 2. Os dados da tabela a seguir são do exercício 1.

Tabela 14 – Dez truncamentos do exercício 1 mais frequentes no *Twitter* e suas frequências no *Web/Dialects*

Truncamento	<i>Twitter</i> (frequência média)	<i>Web/Dialects</i>
bi	3.357,7	4.006

niver	3.090,3	840
refri	1.238,3	466
japa	1.210,0	645
madruga	383,3	640
belê	354,0	13
maraca	300,7	513
oftalmo	260,0	183
profe	177,0	70
flagra	150,3	327

Fonte: autor (2023)

Como visto anteriormente, na busca no *Web/Dialects*, registramos itens com um número entre 10.000 e 100.000 de ocorrências. No *Twitter*, o dado com maior frequência é *bi* com 3.357 ocorrências. Isso possivelmente ocorre pelas limitações que a ferramenta de busca do *Twitter* impõe, principalmente na sua versão gratuita. Mesmo assim, para alguns itens, a frequência foi maior no *Twitter* do que no *Web/Dialects*, como é o caso de *niver*, *refri*, *fapa*, *belê*, *oftalmo* e *profe*. A nossa hipótese para esses casos é que, apesar de o *Web/Dialects* ter dados de sites mais informais, como blogs pessoais, por exemplo, e dados de comentários em sites de notícias, o *Twitter* tem um caráter ainda mais informal e mais próximo da fala. Por isso, alguns dados podem ser mais facilmente encontrados e ter frequências maiores no *Twitter*.

A tabela a seguir compara as ocorrências dos dados do exercício 2 no *Twitter* e no *Web/Dialects*.

Tabela 15 – Dez truncamentos do exercício 2 mais frequentes no *Twitter* e suas frequências no *Web/Dialects*

Truncamento	<i>Twitter</i> (frequência média)	<i>Web/Dialects</i>
fute/futi	1.296	10
		0 (juvenil) 6 (juventude)
juve/juvi	662,3	95 (Juventus)
responosa	619	334
odonto	613	158
neura	348,7	566
bisa	244	128
carna	237,7	4
traveco	216	214
tranks	130,3	38

preju	117,7	21
-------	-------	----

Fonte: autor (2023)

No exercício 2 do experimento aplicado por Heineck (2018), os participantes eram apresentados a palavras do português e convidados a produzirem uma forma truncada a partir destas palavras. Alguns dados que surgiram neste experimento não haviam sido registrados ainda por outros pesquisadores. Mesmo assim, na busca no *Twitter*, dados como *fute/futi* e *juve/juvi* tiveram um número de ocorrências alto em relação aos outros dados. No entanto, na busca *Web/Dialects* obtivemos um resultado diferente. Com exceção de *neura*, em que o número de ocorrências foi maior no *Web/Dialects*, e de *traveco*, em que o número de ocorrências foi muito próximo, os dados tiveram uma frequência menor do que no *Twitter*. Assim como nos casos do exercício 1, acreditamos que esses dados podem ter um caráter bastante informal e, por isso, o *Twitter* seria um espaço de maior uso desse itens.

6.2.3 Conclusões gerais do estudo sobre frequência no Corpus do Português

A pesquisa realizada no Corpus do Português – subcorpus *Web/Dialects* teve por objetivo ampliar o estudo sobre frequência realizado no *Twitter* e colaborar na composição de um *corpus* de dados de truncamento robusto e completo no que diz respeito a suas informações acerca de cada item. Por isso, a busca foi feita com todos os dados do nosso *corpus*.

Mesmo com um *corpus* com uma quantidade considerável de dados como o *Web/Dialects* (cerca de 1 bilhão de palavras), ainda é difícil realizar a busca de dados de truncamento por ser um *corpus* de língua escrita e que lida com alguns textos formais. Desse modo, é esperado que alguns dados de truncamento tenham 0 ocorrências. Em contrapartida, a busca no *Web/Dialects* ampliou a margem de frequência que havíamos encontrado no *Twitter* e registramos itens com até 100 mil ocorrências.

A análise dos itens truncados mais frequentes revelou que eles geralmente têm palavras-base também frequentes. O contrário, no entanto, não se confirma: para algumas das palavras-base mais frequentes do nosso *corpus*, suas respectivas formas truncadas tiveram ocorrências baixas ou até zero ocorrências. Podemos inferir, dessa forma, que truncamentos frequentes parecem ter origem em palavras também frequentes, mas nem toda palavra muito frequente irá gerar truncamentos frequentes.

Na comparação com os dados de Heineck (2021) verificamos que alguns dos dados mais frequentes no *Twitter* não eram frequentes no *Web/Dialects*. A nossa hipótese é a de que o

Twitter é um ambiente mais informal e mais propício para o uso de formas como o truncamento do que os sites que compõe o banco de dados do *Web/Dialects*.

PARTE 3
INVESTIGAÇÃO EXPERIMENTAL

7 TAREFA DE DECISÃO LEXICAL COM *PRIMING*

Nesta etapa da pesquisa, temos por objetivo realizar uma tarefa de decisão lexical com *priming* de modo a avaliar o processamento morfológico do processo do truncamento pelos participantes. A tarefa de decisão lexical com *priming* é, na verdade, a junção de dois protocolos distintos, mas que podem atuar de forma conjunta.

De acordo com Pinto (2017, p. 53), na tarefa de decisão lexical, “os participantes têm de produzir um juízo sobre uma combinação de letras, reconhecendo-a como uma palavra, ou não, tão depressa e acertadamente quanto possível”

No paradigma experimental do *priming*, o processamento de um estímulo apresentado inicialmente (*prime*) pode ou não influenciar o processamento do estímulo apresentado subsequentemente (alvo). O paradigma de *priming* apresenta duas características básicas:

- a presença de um estímulo-*prime* e de um estímulo-alvo;
- o *prime* tem de apresentar uma partilha de traços com o alvo;

Esta técnica assenta no pressuposto de que o *prime* exerce uma influência sobre a resposta ou julgamento de um alvo, funcionando como sua pré-ativação, aumentando a sensibilidade ao alvo. (PINTO 2017, p. 72).

Neste experimento, portanto, iremos reunir a tarefa de decisão lexical com o efeito de *priming*, de modo que os participantes tenham que tomar uma decisão a respeito de uma palavra que lhes será apresentada e terão também a presença de um estímulo-*prime*.

7.1 SELEÇÃO DOS DADOS

Para a seleção dos dados utilizados na tarefa de decisão lexical com *priming*, consideramos os critérios fonológicos e morfológicos utilizados na organização do nosso *corpus*, e buscamos incluir exemplares de todas as combinações do *corpus* no experimento. Incluímos, por exemplo, tanto dados de truncamentos dissílabos e paroxítonos, de base dissílaba e oxítona, quanto de truncamentos dissílabos e oxítonos, de base trissílaba e oxítona. Do mesmo modo, foram também incluídos dados de palavras-base compostas de duas palavras morfológicas (*odontologia*). Um quadro com os dados selecionados para o experimento pode ser encontrado no Anexo 10.4.

O experimento é composto, também, por não palavras que servem como formas distratoras. As não palavras são formas que não seguem padrões da língua portuguesa e tinham por função impedir que os participantes identificassem o objetivo do teste. Elas foram

produzidas no site Random Word Generator¹⁰ e foram selecionadas pela autora. A escolha em utilizar não palavras como itens distratores se deu pelo entendimento do que os participantes levariam em consideração no julgamento do que é ou não uma palavra do português. Considerando que o processo de formação dos truncamentos se dá respeitando condições de boa formação da língua, mesmo que o participante não conhecesse aquela forma truncada, poderia identificá-la como uma palavra possível do português, principalmente no contraste com a ‘má-formação’ das não palavras. Realizamos um experimento piloto em que incluímos pseudopalavras e não-palavras como distratoras e o índice de acertos foi bastante baixo. Acreditamos que, como utilizamos dados de truncamento pouco frequentes, os participantes não diferenciavam as pseudopalavras dos truncamentos e consideravam ambos como palavras não possíveis do português. Este foi outro fator que contribuiu para a escolha de não palavras como itens distratores no experimento.

A fim de não tornar o experimento tão longo e cansativo para os participantes, os dados foram divididos em três. Cada lista compôs uma tarefa de decisão lexical com *priming* diferente e os participantes tiveram que responder a apenas uma das listas cada. Cada participante tinha, portanto, acesso a 1/3 dos itens, totalizando 30 pares experimentais.

7.2 CONDIÇÕES DO EXPERIMENTO

Como mencionado anteriormente, realizamos um experimento online no qual os participantes foram convidados a realizar uma tarefa de decisão lexical com *priming*. Este experimento foi hospedado no site PsyToolkit (Versão 3.4.4) e foi respondido de forma remota, ou seja, cada participante estava utilizando seu próprio computador em sua residência ou local de trabalho. O experimento é composto pelas seguintes etapas:

1. Ao acessar o link do experimento, o participante tem acesso a uma página inicial com as informações sobre o experimento, bem como o Termo de Consentimento Livre e Esclarecido (que pode ser consultado na seção “Anexos”). Após a leitura desta página, o participante precisa confirmar que compreendeu as condições da pesquisa e concordar em participar da pesquisa de forma voluntária para dar sequência ao experimento.
2. Na sequência, o participante responde a um questionário social, com perguntas acerca da sua idade, gênero, escolaridade e estado. É importante ressaltar que a identidade do participante permanece anônima durante todo o experimento. O nome do participante

¹⁰ Disponível em <https://randomwordgenerator.com/fake-word.php>

não é solicitado e o site em que o experimento está hospedado não armazena dados que possam identificar o participante.

3. Após as perguntas iniciais, o participante é apresentado às instruções para realizar a tarefa. As instruções informam que, na tarefa, o participante terá que decidir se a palavra apresentada é ou não é uma palavra possível de ser usada no português de forma rápida e intuitiva. Se o participante entender que a palavra é possível de ser usada no português, deverá apertar a tecla P do teclado do computador. Se entender que ela não é possível de ser usada no português, deve apertar a tecla Q.
4. O participante é informado que fará, primeiramente, uma rodada de treino. Nesta rodada, ele deverá fazer o julgamento de oito palavras.
5. Após a etapa de treino, o participante realiza a mesma tarefa com mais dados. Nesta etapa, ele será apresentado a 30 itens de modo aleatório: 20 palavras e 10 itens distratores (não palavras).

A tarefa dos participantes é decidir se a palavra que lhes é apresentada é uma palavra possível de ser usada no português ou não. Isso é o que caracteriza a tarefa de **decisão lexical**. O efeito de *priming* é a ativação de uma estrutura em função de outra estrutura previamente apresentada. Essa ativação é identificada a partir do tempo de reação do participante e do seu índice de acerto. Neste experimento, a estrutura que é apresentada previamente, chamada de *prime*, é a palavra-base (ex.: fotografia, professor) e o alvo é o truncamento (ex.: foto, profe). A hipótese a ser testada aqui é a de que palavras relacionadas morfologicamente – como é o caso da palavra-base e da sua forma truncada – têm menores tempos de reação do que palavras com outros tipos de relações linguísticas.

Na sequência temporal do experimento, a ordem dos elementos era a seguinte: em um primeiro momento, surgia um ponto de fixação em formato de cruz ‘+’ no centro da tela por 500 ms para fixação do olhar do participante; logo após, surgiam o *prime* com tempo de exposição de 50 ms; na sequência, o alvo, que ficava na tela até a decisão do participante ou por, no máximo, 5000 ms. O *prime* e o alvo eram exibidos em letras maiúsculas, fonte Arial, tamanho 40, também no centro da tela.

Conforme explicitado na seção anterior, os dados foram selecionados de acordo com os critérios fonológicos e morfológicos utilizados na organização do nosso *corpus*. Com esses critérios, visamos analisar se o tempo de reação é menor quando a palavra-base apresentada no *prime* é relacionada ao truncamento apresentada no alvo e se fatores como o acento e o tamanho

das palavras-base e dos truncamentos exerce alguma influência no processamento desses dados pelos participantes. A tabela a seguir apresenta as condições consideradas na escolha dos dados e construção do experimento.

Tabela 16 – Condições para a seleção de dados para a tarefa de decisão lexical com *priming*

<i>Prime</i>	<i>Alvo</i>
Palavra-base trissílaba oxítone	truncamento dissílabo oxítone
Palavra-base trissílaba oxítone	truncamento dissílabo paroxítone
Palavra-base trissílaba oxítone	truncamento trissílabo paroxítone
Palavra-base trissílaba paroxítone	truncamento dissílabo oxítone
Palavra-base trissílaba paroxítone	truncamento dissílabo paroxítone
Palavra-base polissílaba oxítone	truncamento dissílabo oxítone
Palavra-base polissílaba oxítone	truncamento trissílabo paroxítone
Palavra-base polissílaba oxítone	truncamento dissílabo paroxítone
Palavra-base polissílaba paroxítone	truncamento dissílabo oxítone
Palavra-base polissílaba paroxítone	truncamento trissílabo oxítone
Palavra-base polissílaba paroxítone	truncamento dissílabo paroxítone
Palavra-base composta 2 palavras morfológicas	truncamento

Fonte: autor (2023)

7.3 PARTICIPANTES

O experimento foi divulgado em redes sociais (*Facebook, Instagram, WhatsApp, etc.*) e entre os contatos da pesquisadora via e-mail, na forma de lista oculta. A divulgação foi feita através de um convite para a participação da pesquisa (Anexo 10.2), no qual foram explicados os objetivos gerais do estudo. A participação era voluntária e não havia custos aos participantes. Para participar da pesquisa, era necessário ser brasileiro, falante de português brasileiro, ser destro, ter acesso a um computador ou laptop e ser maior de idade (acima de 18 anos de idade). É importante que o participante seja destro, de modo a termos o controle de que todas as

respostas afirmativas (clicar na letra P, que está situada do lado direito do teclado) sejam dadas com a mão dominante do participante, num tempo menos destoante entre os respondentes.

Com base na metodologia proposta por Ulrich (2021), em um experimento de decisão lexical com *priming* em seu estudo sobre sufixos avaliativos, os dados de truncamento de nosso experimento foram divididos em três listas e, portanto, cada lista compôs uma tarefa de decisão lexical com *priming*. Cada tarefa esteve disponível de forma virtual por 72h, em consonância com a metodologia proposta por Schwindt (2021) e Schwindt & Abaurre (2022).

Neste experimento, tivemos um total de 62 participantes: 27 participantes na lista 1, 16 participantes na lista 2 e 19 participantes na lista 3. A idade dos participantes variou entre 18 e 80 anos de idade (média de 35,7 anos). O conjunto de participantes conta com 46 mulheres (74,19%) e 16 homens (25,81%), nenhum participante selecionou a opção “outro” para gênero. Na variável escolaridade, apenas um participante disse cursar ou já ter concluído o Ensino Médio (1,61%), 19 participantes cursam ou já concluíram algum curso superior (30,64%) e 42 participantes estão cursando ou já concluíram algum curso de pós-graduação (67,74%).

Na tabela a seguir, temos o perfil dos respondentes por lista.

Tabela 17 – Perfil dos respondentes da tarefa de decisão lexical com *priming* por lista

Variável	Lista 1 27 informantes	Lista 2 16 informantes	Lista 3 19 informantes
Idade			
Média de idade	38,56 anos	35,06 anos	32,16 anos
Amplitude	20 a 80 anos	18 a 61 anos	21 a 55 anos
Gênero	masculino 12 feminino 15	masculino 1 feminino 15	masculino 3 feminino 16
Escolaridade	ensino médio 0 ensino superior 6 pós-graduação 22	ensino médio 1 ensino superior 5 pós-graduação 10	ensino médio 0 ensino superior 8 pós-graduação 8

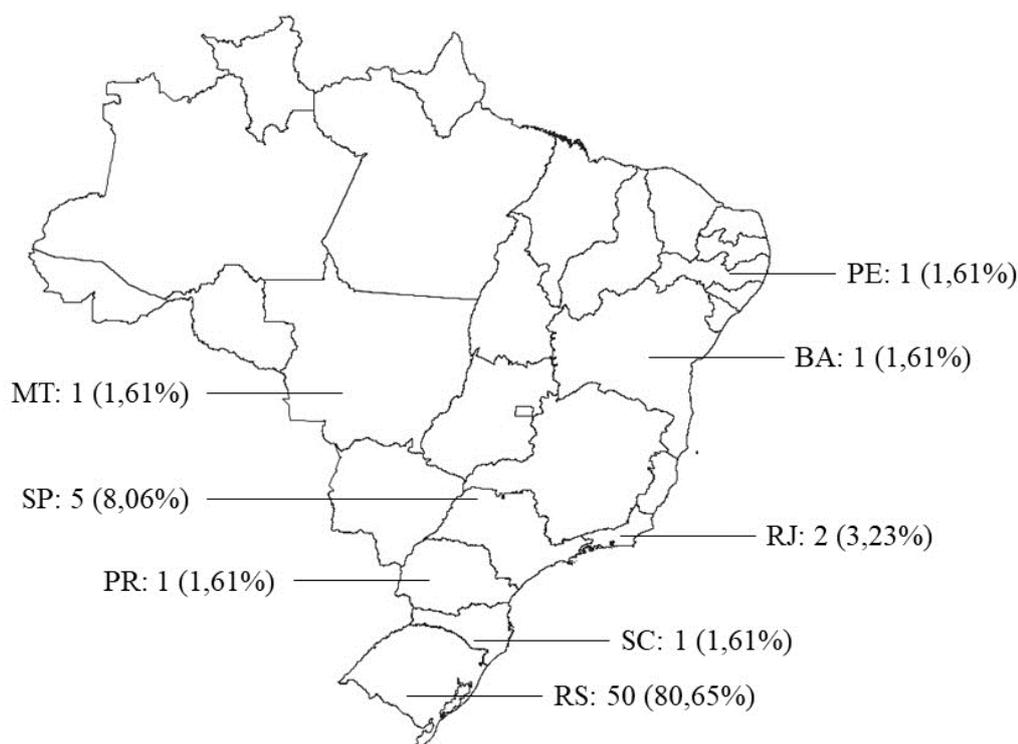
Fonte: autor (2023)

Entre os participantes que declararam ter nível superior ou pós-graduação (61/62), os cursos de formação são bastante heterogêneos. Por proximidade com a pesquisadora, muitos informantes são do curso de Letras (27/62 – 43,55%). Os demais estão distribuídos entre os cursos de Engenharias (6), Direito (6), Pedagogia (4), Administração (4), Arquitetura (2), Ciências Contábeis (2),

Educação Física (2), História (1), Medicina (1), Educação Especial (1), Fisioterapia (1) e Artes Visuais (1), Ciências da Computação (1), Analista de Sistemas (1), Geografia (1).

A maioria dos participantes reside do Rio Grande do Sul (50/62 – 80,65%). Os demais são naturais de outros estados do Brasil.

Figura 8 – Distribuição dos informantes por estados brasileiros na tarefa de decisão lexical com *priming*



Fonte: autor (2023)

É importante destacar, que, como grande parte dos nossos participantes são do Rio Grande do Sul, nossos resultados são predominantemente característicos do dialeto gaúcho.

7.4 ANÁLISE DOS DADOS

Uma vez concluída a coleta, os dados foram codificados em tabela do *software* Excel. Para a análise estatística descritiva e inferencial dos nossos dados, utilizamos o pacote estatístico R 4.3.0, interface RStudio versão 2023.03.1 (R CORE TEAM, 2023). Para a construção dos gráficos, foi utilizado o pacote ggplot2 (WHICKAM, 2016).

As variáveis resposta analisadas são (i) o índice de acertos sobre ser ou não uma palavra do português e (ii) o tempo de reação para a tarefa de decisão lexical (milissegundos).

As variáveis previsoras são:

- (i) a tonicidade da palavra-base: se é oxítônica (*futebol*) ou paroxítônica (*beleza*);
- (ii) a tonicidade do truncamento: se é oxítono (*profí*) ou paroxítono (*jápa*);
- (iii) o número de sílabas da palavra-base: se é trissílaba (*cafuné*) ou possui quatro ou mais sílabas (*adolescente*);
- (iv) o número de sílabas do truncamento: se é dissílaba (*fute*) ou trissílaba (*capita*);
- (v) se a palavra-base é composta por duas palavras morfológicas (*paleontologia*)¹¹;
- (vi) relação da palavra-base com o truncamento: se há relação morfológica entre o *prime* (palavra-base) e o alvo (truncamento).

Nas próximas seções, iremos analisar os resultados do experimento de com base nas variáveis dependentes e independentes listadas.

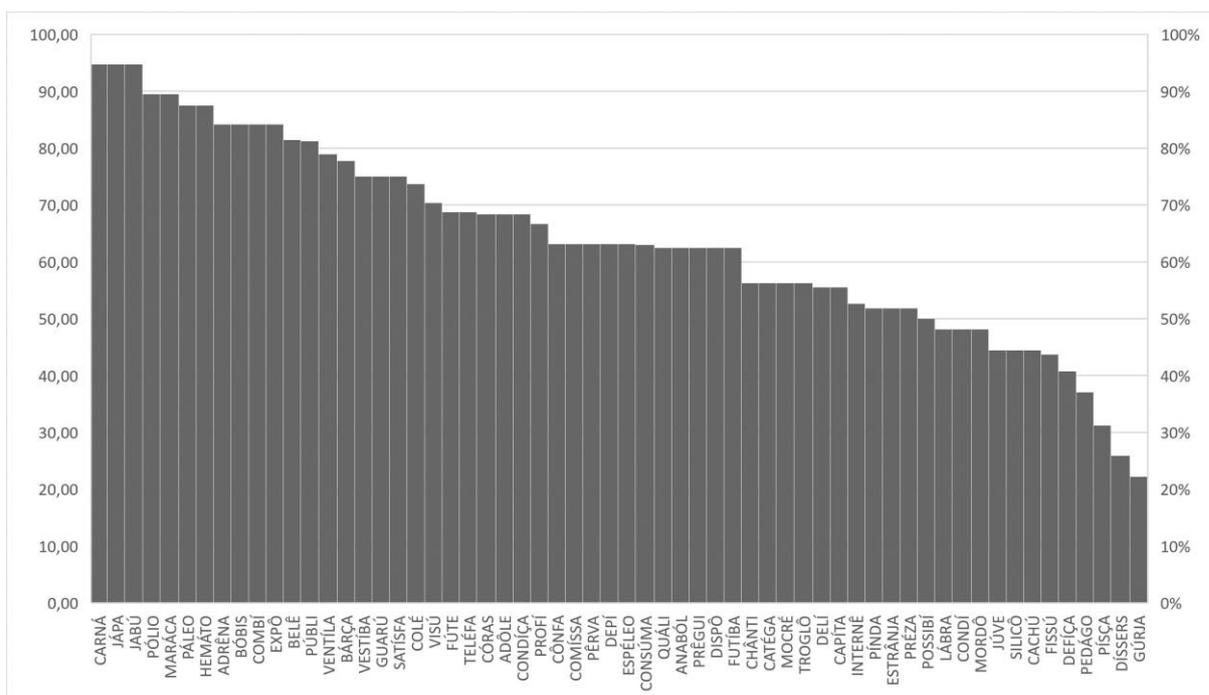
7.4.1 Índice de acerto

O nosso total de dados (1860) é dividido entre palavras reais (1240) e não palavras (620). A proporção do total de acertos em palavras reais é de 773/1240 (62,34%) e a proporção de falsos alarmes – respostas positivas para estímulos inexistentes – é de 39/620 (6,29%). Para a análise do tempo de reação, excluimos as respostas incorretas, ou seja, quando uma palavra real foi lida como não palavra e vice e versa e as não palavras e restaram apenas os dados de acerto nas palavras reais (773).

Olhamos, ainda, para cada um dos estímulos para identificarmos aqueles que tiveram índices de erro maiores. O gráfico abaixo apresenta uma escala decrescente dos índices de acerto dos dados.

¹¹ É importante destacar que o limite morfológico, que delimita as raízes eruditas, coincide, nos dados selecionados para o experimento, com o limite da palavra fonológica. Por esse motivo, essa variável é analisada juntamente com as variáveis relacionadas ao acento.

Gráfico 7 – Índice de acerto dos dados da tarefa de decisão lexical com *priming*



Fonte: autor (2023)

As 5 palavras truncadas com os menores índices de acerto foram *gúrja* (22,22%), *díssers* (25,93%), *pisça* (31,25%), *pedágo* (37,05%) e *defiça* (40,74%). Desses itens, *gúrja*, *díssers* e *pisça* tinham como *prime* a sua própria palavra-base (*gorjeta*, *dissertação* e *piscina*). *Pedágo* e *defiça* tinham como *prime* palavras-base não relacionadas a eles. Acreditamos que esse baixo índice de acerto se deva a familiaridade e frequência destes itens. Os truncamentos que apresentaram mais de 90% de índice de acerto foram *carná*, *jápa* e *jabú* – todos com um índice de 94,74% de acerto.

7.4.1.1 Relação entre *prime* e alvo

Para a análise do papel das variáveis independentes no índice de acerto, excluimos os dados de não palavras. Na tabela abaixo, observamos a distribuição do índice de acerto em relação a variável *relação entre prime e alvo*.

Tabela 18 – Distribuição do índice de acerto em relação à variável *relação entre prime e alvo*

	<i>Prime e alvo relacionados</i>	<i>Prime e alvo não relacionados</i>
Acerto	583 62,69%	190 61,29%
Erro	337 36,24%	119 38,39%
Sem resposta	10 1,08%	1 0,32%

Fonte: autor (2023)

Na distribuição dos dados, observamos índices de acerto próximos entre os itens relacionados e os não relacionados. O teste de qui-quadrado confirmou esta percepção e mostrou não haver diferença significativa entre as duas condições ($X^2(2) = 1.8501$, $p = 0.3965$).

7.4.1.2 Características morfofonológicas da palavra-base

A tabela abaixo apresenta a distribuição dos dados e o índice de acerto de cada item da variável *características morfofonológicas da palavra-base*.

Tabela 19 – Distribuição do índice de acerto em relação à variável *características morfofonológicas da palavra-base*

	2 palavras morfológicas	Polissílaba oxítone	Polissílaba paroxítone	Trissílaba oxítone	Trissílaba paroxítone
Acerto	67 69,07%	143 64,71%	272 61,12%	188 63,95%	103 56,28%
Erro	29 29,90%	73 33,03%	173 38,88%	102 34,69%	79 43,17%
Sem resposta	1 1,03%	5 2,26%	0 0%	4 1,36%	1 0,55%

Fonte: autor (2023)

Na análise da distribuição dos dados, parece não haver diferença importante entre os grupos analisados. Assim como na variável anterior, o teste de qui-quadrado mostrou não haver diferença significativa entre as variáveis ($X^2(2) = 7.1293$, $p = 0.1292$).

7.4.1.3 Características morfofonológicas do truncamento

A distribuição dos dados e o índice de acerto para cada item da variável *características morfofonológicas do truncamento* pode ser observada na tabela abaixo.

Tabela 20 – Distribuição do índice de acerto em relação à variável *características morfofonológicas do truncamento*

	Base com 2 palavras morfológicas	Dissílabo oxítono	Dissílabo paroxítono	Trissílabo oxítono	Trissílabo paroxítono
Acerto	67 69,07%	253 66,06%	203 57,51%	40 51,28%	210 63,83%
Erro	29 29,90%	129 33,68%	144 40,79%	38 48,71%	116 35,26%
Sem resposta	1 1,03%	1 0,26%	6 1,70%	0 0%	3 0,91%

Fonte: autor (2023)

O teste de qui-quadrado indicou uma diferença significativa entre as variáveis ($X^2(2) = 11.479$, $p = 0.02168$). Essa diferença parecia estar nas variáveis truncamentos dissílabos paroxítonos e truncamentos trissílabos oxítonos. No entanto, ao realizar uma análise dos resíduos padronizados ajustados, essa diferença não se confirmou. Portanto, parece não haver diferença significativa real entre as variáveis.

7.4.2 Tempo de reação

Como já comentado no início deste capítulo, para a análise do tempo de reação, foram excluídos os dados cujas respostas foram incorretas ou não houve resposta pelo participante (o tempo de resposta excedeu 5.000ms e o item foi substituído pelo próximo da lista). Os 773 dados experimentais resultantes apresentam mediana igual a 1.026ms.

Nas seções a seguir, iremos analisar os papéis das variáveis *relação entre prime e alvo*, *características morfofonológicas da base* e *características morfofonológicas do truncamento* em relação ao tempo de reação.

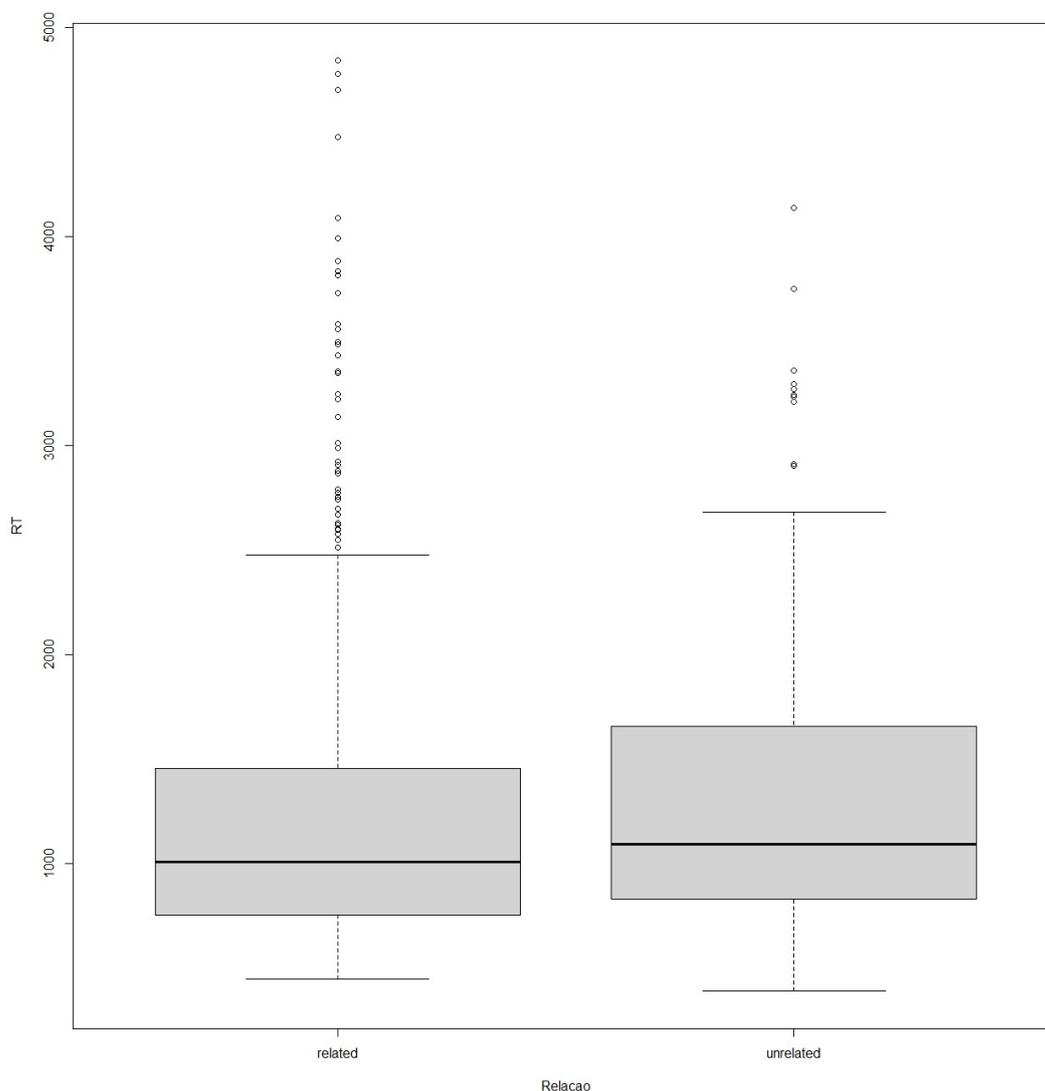
7.4.2.1 Relação entre *prime* e alvo

Essa análise tem por objetivo analisar o papel da relação entre *prime* e alvo na variável dependente tempo de reação, ou seja, pretendemos analisar se o fato de o *prime* e o alvo estarem relacionados (*cachoeira ~ cachú*) ou não (*ventilador ~ bóbis*) influencia no tempo que o participante leva para decidir se reconhece o item como palavra ou não e clicar a tecla adequada referente à sua resposta. Tínhamos, como hipótese inicial, que o tempo de reação seria menor quando o *prime* apresentasse relação com o alvo.

Realizamos um teste Shapiro-Wilk para testar a normalidade dos nossos dados e os resultados apontam para uma distribuição não normal (*prime* e alvo relacionados: $W = 0.7935$, $p < 0,01$; *prime* e alvo não relacionados: $W = 0.8537$, $p < 0,01$). Portanto, os testes realizados nessa análise foram testes não-paramétricos, adequados a esse tipo de distribuição de dados.

O gráfico mostra que os dados em que há relação entre *prime* e alvo (*related*: mediana 966ms/IQR 649ms) apresentam padrões distintos daqueles em que não há essa relação (*unrelated*: mediana 1218ms/IQR 1002ms).

Gráfico 8 – Distribuição dos dados da variável relação entre *prime* e alvo em relação ao tempo de reação



Fonte: autor (2023)

O teste de Wilcoxon apontou para uma diferença significativa ($W = 48.686$, $p = 0.0122$) na mediana do tempo de reação das duas realidades investigadas. Dessa forma, confirmamos a hipótese sobre o papel da relação entre *prime* e alvo no tempo de reação: o tempo de reação é menor quando há relação entre *prime* e alvo, ou seja, quando o *prime* é uma palavra-base cujo alvo é sua respectiva palavra truncada (*cachoeira* ~ *cahú*). Esse resultado aponta para o fato de que os falantes reconhecem estruturas morfológicamente relacionadas, pois respondem mais rapidamente a esse tipo de informação.

7.4.2.2 Características morfofonológicas da palavra-base

O teste Shapiro-Wilk por grupos atestou, novamente, a não normalidade dos dados e, portanto, realizamos um teste não-paramétrico para a análise da influência das características morfofonológicas da palavra-base no tempo de reação.

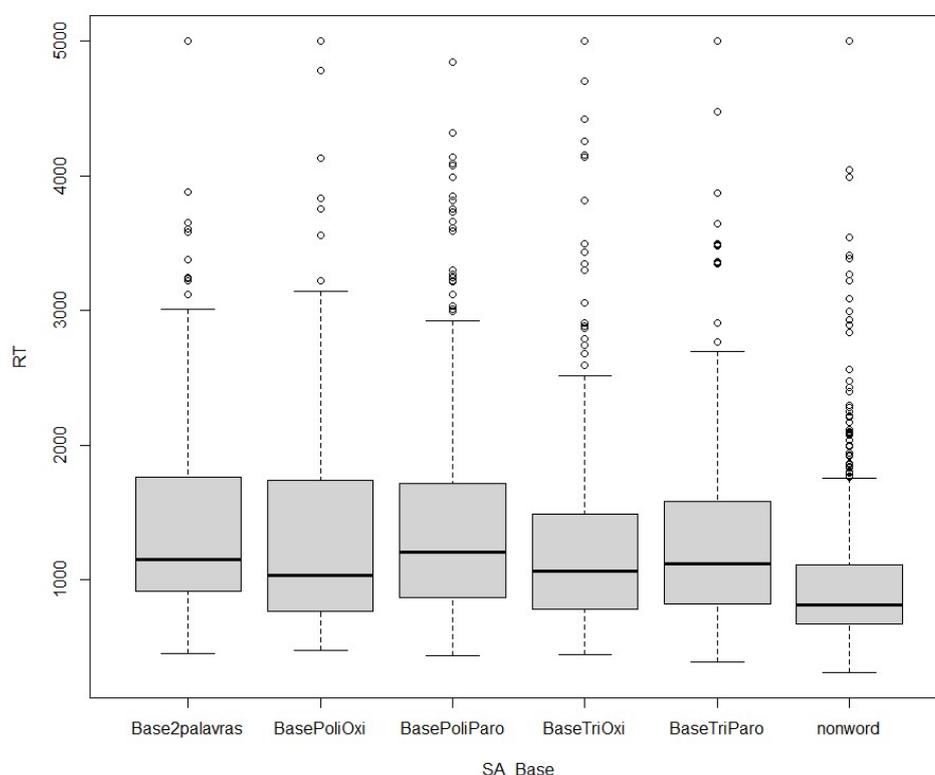
Ao analisar o gráfico e a tabela de distribuição do tempo de reação entre as realidades investigadas não parece apresentar diferenças muito grandes.

Tabela 21 – Medianas do tempo de reação em relação à variável *características morfofonológicas da palavra-base*

Palavra-base	Mediana	IQR
2 palavras morfológicas	1.052	608
Polissílaba oxítone	938	860
Polissílaba paroxítone	1.090	714
Trissílaba oxítone	992	675
Trissílaba paroxítone	1.004	690

Fonte: autor (2023)

Gráfico 9 - Distribuição dos dados da variável *características morfofonológicas da palavra-base* em relação ao tempo de reação



Fonte: autor (2023)

Confirmando o que vimos na distribuição dos dados, o teste Kruskal-Wallis apontou que não há diferença significativa entre as medianas dos grupos analisados ($X^2(2) = 5.471, p < 0.2423$). Portanto, as características morfofonológicas analisadas da palavra-base não apresentaram, neste recorte amostral, papel significativo sobre o tempo de reação dos participantes.

7.4.2.3 Características morfofonológicas do truncamento

Em relação à análise da relação entre as características morfofonológicas do truncamento e o tempo de reação, mais uma vez identificamos não normalidade ($p < 0,01$) dos dados e, portanto, aplicamos testes estatísticos não paramétricos.

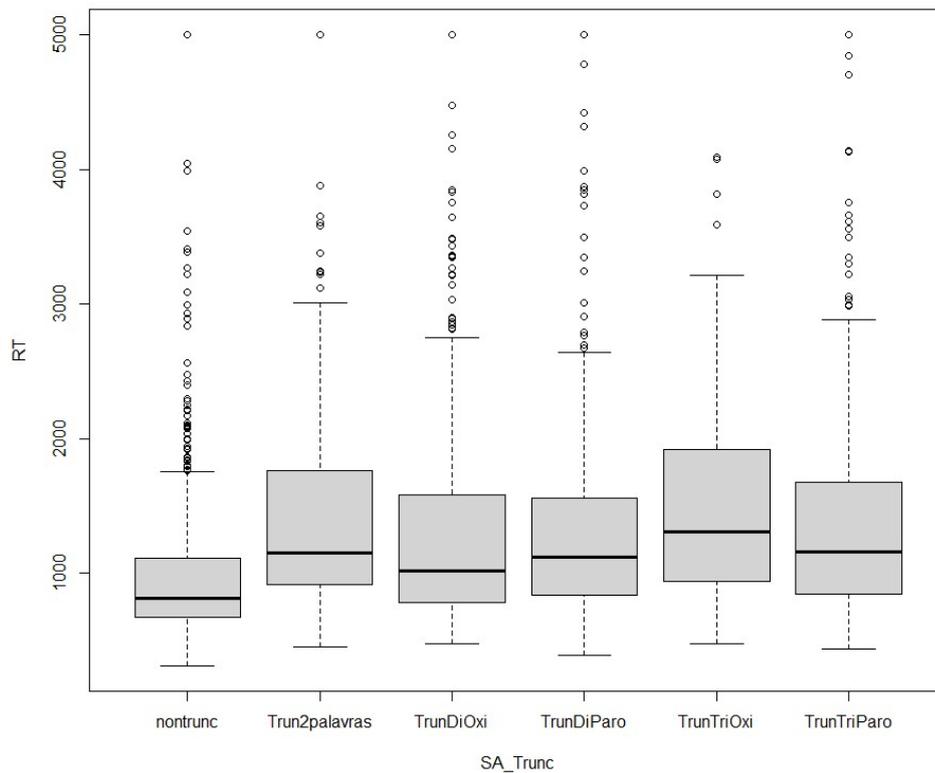
Assim como nas características morfofonológicas da base, aqui, também não observamos grandes diferenças na distribuição dos dados, como vemos na tabela e no gráfico abaixo.

Tabela 22 – Medianas do tempo de reação em relação à variável *características morfofonológicas do truncamento*

Truncamento	Mediana	IQR
Base com 2 palavras fonológicas	1.052	608
Dissílabo oxítono	905	727
Dissílabo paroxítono	1.071	702
Trissílabo oxítono	1.032	742
Trissílabo paroxítono	1.082	770

Fonte: autor (2023)

Gráfico 10 - Distribuição dos dados da variável *características morfofonológicas do truncamento* em relação ao tempo de reação



Fonte: autor (2023)

O teste Kruskal_Wallis confirmou nossa percepção de que não há diferença significativa entre as medianas dos itens observados nessa variável. Dessa forma, não há influência das características morfofonológicas do truncamento no tempo de reação dos participantes.

7.5 Conclusões gerais da tarefa de decisão lexical com *priming*

O índice geral de acerto não foi tão alto quanto esperado (72,69%), principalmente quando excluímos as não palavras e olhamos apenas para as palavras reais (62,34%). Entendemos que isso pode ter ocorrido devido à escolha dos itens experimentais. Optamos por dados pouco frequentes, e que acreditamos serem pouco familiares aos participantes, de modo a encobrir o objetivo do experimento.

Para uma possível continuidade do estudo, pode ser interessante realizar um teste de familiaridade para que possamos identificar dados mais familiares e aceitos pelos participantes. O teste pode consistir em uma tarefa de associação de palavras, conforme propõe Ulrich (2021) em seu estudo sobre sufixos avaliativos, ou um teste de reconhecimento lexical, conforme Minussi e Vilalva (2020) numa pesquisa sobre *blends*.

As variáveis observadas – relação entre *prime* e alvo, características morfofonológicas da palavra-base e características morfofonológicas do truncamento – não apresentaram diferenças significativas quanto ao índice de acerto e, portanto, não parecem exercer influência na escolha de resposta dos participantes.

Quanto ao tempo de reação, confirmamos a nossa hipótese inicial: o tempo de reação é menor quando há relação morfológica entre *prime* e alvo, ou seja, quando havia relação entre palavra-base e o truncamento, o tempo de resposta dos participantes era menor. Isso parece confirmar que o processamento da forma truncada é mais rápido quando o falante o relaciona com sua palavra-base.

No que diz respeito às características morfofonológicas tanto da palavra-base quanto do truncamento, os testes estatísticos não apresentaram diferenças significativas nas medianas destas variáveis em relação ao tempo de reação. Desse modo, elas não parecem exercer influência no tempo de processamento dos itens analisados.

Por fim, é importante salientar a não normalidade de dados, o que implicou o emprego de testes estatísticos não paramétricos. Para a continuidade deste estudo, desejamos ampliar a amostra e fazer uso de dados mais familiares pelos participantes, de modo a não excluir tantos dados com respostas erradas. Dessa forma, acreditamos que possamos ter uma distribuição normal dos dados e, conseqüentemente, aplicar testes estatísticos paramétricos em nossa análise. Nessa perspectiva, não se descarta, inclusive, encontrar condicionamento das variáveis morfofonológicas, para as quais não se identificou papel significativo neste estudo.

8 CONSIDERAÇÕES FINAIS

Este trabalho tinha como objetivo a descrição do fenômeno do truncamento no português brasileiro, delimitação de suas regularidades e a contribuição da discussão acerca do fenômeno feita em estudos anteriores. Iremos retomar, a seguir, as questões que nortearam nosso trabalho e que foram apresentadas na introdução, respondendo-as, quando possível, a partir dos nossos resultados e conclusões.

a) De que forma fatores como o número de sílabas e acento influenciam na formação de truncamentos?

Nos estudos de Heineck (2018), identificamos um possível papel do tamanho da palavra-base na formação de truncamentos. Por esse motivo, selecionamos dados obtidos no teste *online* aplicado por Heineck (2018) e realizamos uma nova análise dos dados com foco no papel do número de sílabas. Foram codificados tanto o tamanho do dado apresentado ao informante na pergunta dos exercícios, quanto o dado da resposta do informante. Realizamos análises descritivas e inferenciais dos dados no *software* SPSS.

Na análise, ainda inicial, realizada neste trabalho, identificamos que quando havia um aumento no número de sílabas da palavra-base, as formas truncadas tendiam a ser maiores. Além de confirmar os resultados de Heineck (2018), esta análise confirma os resultados de Martini (2010), que identifica um padrão na formação de truncamentos: palavras dissílabas geralmente formam truncamentos também dissílabos, palavras trissílabas geralmente formam truncamentos dissílabos e palavras polissílabas tendem a formar truncamentos trissílabos. Para obter resultados mais precisos, é preciso ampliar a análise, visto que os dados utilizados neste estudo não apresentaram uma distribuição normal.

Os dados de truncamento levantados a partir de estudos anteriores sobre o fenômeno ((ARAÚJO, 2002; VILELA, GODOY & SILVA, 2006; SCHER, 2011; BELCHOR, 2014; SCHER, 2018), dos resultados obtidos em Heineck (2018) e do registro de dados de redes sociais, programas de televisão, etc., foram reorganizados em uma tabela *Excel* e compõem o nosso *corpus* com um total de 239 dados. Estes dados foram decodificados a partir de critérios fonológicos e morfológicos, assim como havia sido feito por Heineck (2018), e foram acrescentadas anotações em relação a frequência da forma truncada e de sua palavra-base, bem como exemplos de uso de cada dado.

Quando observamos a distribuição dos dados no nosso *corpus* em relação ao número de sílabas, identificamos que se assemelha ao constatado em Heineck (2018), em relação aos resultados dos experimentos. Há uma tendência de o truncamento ser menor que a palavra-base, visto que isso ocorre em 85,36% dos dados do *corpus*. Os dados do *corpus*, assim como os resultados da dissertação, também corroboraram a tendência proposta por Martini (2010), que aponta que, geralmente, palavras-base dissílabas formam truncamentos dissílabos, palavras-base trissílabas formam truncamentos dissílabos e, por fim, palavras-base polissílabas formam truncamentos trissílabos.

A correlação entre o tamanho da palavra-base e o tamanho do truncamento observada na análise estatística inferencial realizada com os dados de Heineck (2018) não parece se confirmar na distribuição dos dados do nosso *corpus*, visto que a maioria dos truncamentos é dissílaba, independentemente do número de sílabas da palavra-base. Há um aumento de dados de truncamentos trissílabos quando a palavra-base é polissílaba, mas isso pode ser devido à tendência de a forma truncada ser menor que a sua base e, por esse motivo, haveria mais dados de truncamentos trissílabos com palavras polissílabas do que com palavras-base trissílabas.

Nos dados do nosso *corpus*, grande parte das formas truncadas é paroxítona, compondo 76,66% do total de dados. Essa distribuição dos dados vai ao encontro do que foi encontrado em Heineck (2018), em que a maioria dos dados era dissílaba e paroxítona. A formação de truncamentos paroxítonos é esperada, visto que esse é o padrão acentual mais regular da língua portuguesa. Não parece haver, portanto, uma relação entre o padrão acentual da palavra-base e de seu truncamento.

b) O truncamento é um processo produtivo de formação de palavras?

Com o principal objetivo de investigar o papel da frequência no processo do truncamento, realizamos uma busca na rede social *Twitter* dos dados obtidos nos exercícios 1 e 2 dos experimentos aplicados por Heineck (2018). Para a busca, foi utilizado o pacote *rtweet* do programa computacional R. Como objetivos específicos desta etapa do trabalho, pretendeu-se investigar se os falantes ainda conseguiam recuperar a palavra-base de truncamentos muito frequentes na língua e se os dados mais frequentes nos resultados de Heineck (2018) eram também os mais frequentes no *Twitter*.

De modo a ampliar a pesquisa realizada no *Twitter*, verificamos a frequência dos itens que compõem o nosso *corpus* de dados de truncamento no Corpus do Português – subcorpus *Web/Dialects*, que possui cerca de 1 bilhão de palavras. O Corpus tem dados de 4 países –

Brasil, Portugal, Angola e Moçambique –, no entanto, a nossa busca restringiu a dados do Brasil.

Na busca realizada no *Twitter*, houve poucos dados com zero ocorrências e, entre os dados mais frequentes, o número de ocorrências variou entre 1.000 e mais de 3.000. Já na busca com dados do Corpus do Português – *Web/Dialects*, mesmo com um corpus com uma grande quantidade de dados, ainda houve alguns itens do nosso *corpus* com zero ocorrências. Isso confirma o caráter informal de uso das formas truncadas, pois o *Web/Dialects* é um *corpus* de língua escrita e, apesar de seus dados serem de sites da internet, alguns têm como origem textos mais formais, como as reportagens jornalísticas. Em contrapartida, a busca no *Web/Dialects* ampliou a margem de frequência que havíamos encontrado no *Twitter* e registramos itens com até 100 mil ocorrências. As buscas parecem apontar para um processo que é produtivo, mesmo que, em relação a outros processos de formação de palavras do português, não seja tão frequente.

- c) A frequência da palavra-base tem influência na produção de formas truncadas? A frequência da forma truncada depende da frequência de sua palavra-base ou é possível que um truncamento seja mais frequente que sua palavra de origem?

Os resultados obtidos na análise de frequência com dados do *Twitter* parecem confirmar que, mesmo quando os truncamentos são muito frequentes na língua, os falantes conseguem recuperar sua palavra-base, pois os truncamentos nos quais os falantes recuperavam com facilidade a base nos experimentos de Heineck (2018) também eram os mais frequentes entre os dados buscados no *Twitter*.

Verificamos que, no exercício 2, no qual os falantes deveriam propor formas truncadas para palavras do português, os dados de truncamento mais frequentes nos experimentos de Heineck (2018) também eram, de um modo geral, os mais frequentes 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 por eles conhecidas.

A partir da análise dos truncamentos e das palavras-base mais frequentes do nosso *corpus* no Corpus do Português, verificamos que os truncamentos mais frequentes tinham como base palavras frequentes, mas algumas das palavras-base mais frequentes desse *corpus* têm formas truncadas com poucas ou até zero ocorrências. Pudemos inferir, nesse sentido, que

truncamentos frequentes parecem ter origem em palavras também frequentes, mas que nem toda palavra muito frequente irá gerar truncamentos frequentes.

Ao comparar os resultados de Heineck (2021) com os resultados da busca no *Web/Dialects*, verificamos que alguns dos dados mais frequentes no *Twitter* não eram frequentes no *Web/Dialects*. Isso pode ocorrer, pois o *Twitter* é um ambiente mais informal e mais propício para o uso de formas como o truncamento do que os sites que compõe o banco de dados do *Web/Dialects*.

d) de que forma os falantes processam as formas truncadas? o processamento é mais rápido e eficiente quando os falantes relacionam o truncamento com sua palavra-base?

De modo a avaliar o processamento morfológico do processo do truncamento pelos falantes de português, realizamos um experimento *online* que consistia em uma tarefa de decisão lexical com *priming*. No que diz respeito à tarefa decisão lexical, os participantes eram expostos a palavras truncadas e a não palavras (itens distradores) e precisavam decidir se os itens eram ou não palavras possíveis do português. O efeito de *priming* se dá a partir da ativação de uma estrutura em função de outra estrutura previamente apresentada. No nosso caso, a estrutura que era apresentada previamente, chamada de *prime*, era a palavra-base (ex.: fotografia, professor) e o alvo era o truncamento (ex.: foto, profe). Pretendíamos testar se palavras com relação morfológica — como é o caso da palavra-base e sua forma truncada — teriam índices de acerto maiores e/ou menores tempo de reação do que palavras não relacionadas.

O índice de acerto da tarefa de decisão lexical com *priming* não foi tão alto quanto o esperado. Acreditamos que isso pode ter ocorrido pela escolha de dados de truncamento pouco frequentes, pois tínhamos o objetivo de não deixar tão claro o objeto de estudo do experimento para os participantes.

Os testes estatísticos mostraram que o índice de acertos não teve diferença significativa em relação às variáveis observadas. Portanto, a relação entre *prime* e alvo, as características morfofonológicas da palavra-base e as características morfofonológicas do truncamento não exerceram influência na escolha de resposta dos participantes.

Quanto ao tempo de reação, a nossa hipótese inicial foi confirmada: o tempo de reação é menor quando há relação morfológica entre *prime* e alvo, ou seja, quando havia relação entre

palavra-base e o truncamento, o tempo de resposta dos participantes era menor. Isso parece confirmar que o processamento da forma truncada é mais rápido quando o falante o relaciona com sua palavra-base.

As características morfofonológicas tanto da palavra-base quanto do truncamento não apresentaram diferenças significativas nas suas medianas em relação ao tempo de reação. Com base nesse resultado, podemos dizer que não exercem influência no tempo de processamento dos itens da amostra investigada.

É importante salientar a não normalidade de dados, o que implicou o emprego de testes estatísticos não paramétricos. Para a continuidade deste estudo, desejamos ampliar a amostra e fazer uso de dados mais familiares para os participantes, de modo a não excluir tantos dados com respostas erradas. Para a seleção desses dados poderíamos realizar testes para atestar familiaridade dos itens, como uma tarefa de associação de palavras, conforme propõe Ulrich (2021) em seu estudo sobre sufixos avaliativos, ou um teste de reconhecimento lexical, conforme sugerem Minussi e Vilalva (2020) em pesquisa sobre *blends*. Dessa forma, acreditamos que possamos ter um índice de acertos maior, uma distribuição normal dos dados e, conseqüentemente, aplicar testes estatísticos paramétricos em nossa análise.

Por fim, este trabalho teve como principal objetivo oferecer uma análise descritiva e empírica do fenômeno do truncamento e contribuir para os estudos e discussões acerca do processo. Identificamos como uma fragilidade desse estudo e possível perspectiva de continuidade uma discussão teórico-analítica que busque uma proposta de modelagem do fenômeno do truncamento em teorias formais e que dialogue com os estudos presentes na literatura em diferentes concepções teóricas.

9 REFERÊNCIAS BIBLIOGRÁFICAS

ALCÂNTARA, C. C. As classes formais do português brasileiro. *Letras de Hoje* 45(1): 5– 15, 2010.

ALCÂNTARA, C. C. As classes formais do português e sua constituição: um estudo à luz da teoria da morfologia distribuída. Tese (Doutorado) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

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.

BASILIO, M. *Teoria lexical*. 3. ed. São Paulo: Ática, 1991.

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

BENUA, L. Identify effects in morphological truncation. In: BECKMAN, J. (Ed.). *Papers in Optimality Theory*, Massachusetts, v.18, n.1, p.77-136, 1995.

BERLCHOR, 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.

DAVIES, Mark. *Corpus do Português: Web/Dialects*, 2016. Disponível em: <http://www.corpusdoportugues.org/web-dial/>.

FIDELHOLTZ, James 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.; VAZQUEZ R. P. Fla x Flu no Maraca: uma análise otimalista do truncamento no português do Brasil. In: SILVA, J. P. (org.) *Questões de morfossintaxe*. v. 8, p. 56-64, Rio de Janeiro, 2004.

GONÇALVES, C. A. V. Interface morfologia-fonologia: teorias, abordagens e tema. *Cadernos de Letras da UFF: Dossiê: Língua em uso*, Niterói, n.47, p. 333-355, 2013.

GONÇALVES, C.A.V. *Introdução à morfologia não linear*. Rio de Janeiro, Publit: 2009.

GONÇALVES, C.A.V. Processos Morfológicos Não concatenativos no português do Brasil: tipos e funções. [Versão revista e ampliada da Comunicação “Processos de redução vocabular: tipos e funções”, apresentada na XVII Jornada de Estudos Lingüísticos do GELNE, Recife, Instituto de Letras da UFPE, set. 1999.

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.

HALLE, M.; MARANTZ, A. Distributed morphology and pieces of inflection. In K. Hale, and S. J. Keyser (Org.), *The View from the Building 20: Essays in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, 111– 176, 1993.

HEINECK, D. Aspectos morfofonológicos do truncamento no português brasileiro. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul, 2018.

HEINECK, D. Truncamento no português brasileiro: um estudo de frequência com dados do Twitter. *Diadorim*, Rio de Janeiro, vol. 23, n. 2, p. 383-400, jul.-dez. 2021.

HUBACK, A. P. A interferência da frequência em fenômenos linguísticos. *Revista Delta*, 29:1, p. 79–94, 2013.

MARTINI, L. D. *Morfologia Prosódica no Português Brasileiro*. Belo Horizonte: Faculdade de Letras/UFMG, 2010.

MCCARTHY, J. & PRINCE, A.S. *Prosodic Morphology*. Amherst: University of Massachusetts and Brandeis University, 1986.

MINUSSI R. D.; VILLALVA A. Reconhecimento e acesso lexical dos blends em Português Europeu e Português Brasileiro. *Todas as Letras* 22.1: 1-14, 2020.

PINTO, C. A. G. O Papel da Estrutura Morfológica nos Processos de Leitura de Palavras. Tese de Doutorado. Universidade de Lisboa, 2017.

R CORE TEAM. “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria, 2023. Disponível em: <<https://www.r-project.org/>> Último acesso em 09 de junho de 2023.

SANDMANN, A. *Morfologia lexical*. São Paulo: Contexto, 1992.

SANTOS, J.B.A. *Truncamento no português do Brasil: acaso ou processo?* Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, 2002.

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, p. 332-353, 2016.

SCHER, A. P. Formas truncadas em português brasileiro e espanhol peninsular: descrição preliminar. *ReVEL*, edição especial n. 5, 2011.

SCHWINDT L.C., ABAURRE, M.B. On the emergence of [n] in the derivation of nasal-final words in brazilian portuguese. *J. of Speech Sci.* Vol 11, 2022.

SCHWINDT, L. C. *O prefixo no Português Brasileiro: análise morfofonológica*. Tese de doutorado. Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul, 2000.

SCHWINDT, L. C. *Underlying representation of [w]-final words in Brazilian Portuguese: evidence from morphological derivation*. *Acta Linguistica Academica*, v. 68, n. 1-2, p. 139-157, 2021.

SELKIRK, E. *The syntax of words*. Cambridge, MA: MIT Press, 1982.

ULRICH, C. W. O estatuto morfológico e prosódico da sufixação avaliativa em português brasileiro. Tese (Doutorado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul, 2021.

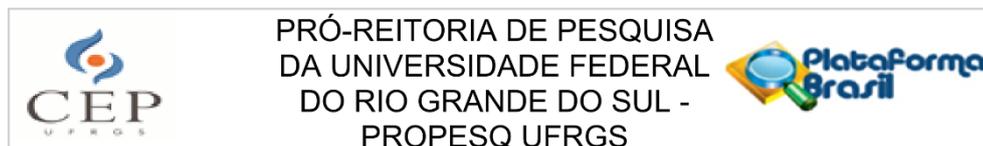
VILELA, A. C.; GODOY, L. & SILVA, T. C. Truncamento no PB - para uma melhor compreensão do fenômeno. 2006.

VILLALVA, A. Morfologia do Português. Lisboa: Universidade Aberta, 2007.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, 2016. Disponível em: <https://ggplot2.tidyverse.org>.

10 ANEXOS

10.1 APROVAÇÃO DA PESQUISA PELO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: INVESTIGAÇÃO EXPERIMENTAL SOBRE O FENÔMENO DO TRUNCAMENTO NO PORTUGUÊS BRASILEIRO

Pesquisador: Luiz Carlos da Silva Schwindt

Área Temática:

Versão: 2

CAAE: 67487023.8.0000.5347

Instituição Proponente: Universidade Federal do Rio Grande do Sul

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 6.015.577

Apresentação do Projeto:

As informações elencadas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram retiradas do documento Informações Básicas da Pesquisa, n.º 2091297, datado e, 08/04/2023, e "Projeto Detalhado", arquivo Projeto_ComiteEtica_DeboraHeineck_atualizado_CEP.

O projeto apresentado, intitulado "INVESTIGAÇÃO EXPERIMENTAL SOBRE O FENÔMENO DO TRUNCAMENTO NO PORTUGUÊS BRASILEIRO" deriva da tese de Doutorado intitulada "Processos de encurtamento vocabular em português brasileiro: análise do truncamento e da hipocorização", que está sendo desenvolvida por Débora Heineck, sob orientação do professor Luiz Carlos da Silva Schwindt (pesquisador responsável), do Programa de Pós-Graduação em Letras da UFRGS.

O projeto tem por objetivo geral investigar e descrever o funcionamento do processo de truncamento em português brasileiro e promover uma discussão teórico-analítica das formas de base e das restrições fonológicas e morfológicas que dão conta do processo, de modo a contribuir para a discussão a respeito do funcionamento deste processo e para o seu entendimento do ponto de vista do seu enquadramento na gramática do português e das línguas naturais de um modo geral. Como objetivos específicos, podemos mencionar: (i) Compor um corpus de dados atuais de

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br

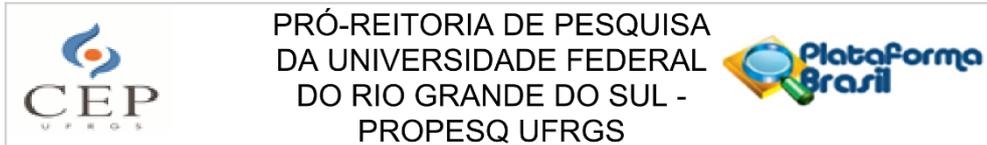
truncamentos que possa contribuir para os corpora de dados deste processo já existentes; (ii) Investigar de que modo se dá o processamento morfológico do truncamento por falantes do português brasileiro, através de um experimento virtual de decisão lexical com priming; (iii) Discutir o papel de fatores fonológicos como a tonicidade e o número de sílabas para a formação de truncamentos; (iv) Investigar o processamento morfológico de truncamentos formados a partir de palavras compostas por duas raízes eruditas (psicologia).

Metodologia:

Será realizado um experimento online no qual os participantes serão convidados a realizar uma tarefa de decisão lexical com priming. Este experimento está hospedado no site PsyToolkit. O experimento é composto pelas seguintes etapas:

1. Ao acessar o link do experimento, o participante tem acesso a uma página inicial com as informações sobre o experimento, bem como o Termo de Consentimento Livre e Esclarecido (que pode ser consultado na seção "Anexos"). Após a leitura desta página, o participante precisa confirmar que compreendeu as condições da pesquisa e concordar em participar da pesquisa de forma voluntária para dar sequência ao experimento.
2. Na sequência, o participante responde a um questionário social, com perguntas acerca da sua idade, gênero, escolaridade e estado. É importante ressaltar que a identidade do participante permanece anônima durante todo o experimento. O nome do participante não é solicitado e o site em que o experimento está hospedado não armazena dados que possam identificar o participante.
3. Após as perguntas iniciais, o participante é apresentado às instruções para realizar a tarefa. As instruções informam que, na tarefa, o participante terá que decidir se conhece a palavra apresentada ou não de forma rápida e intuitiva. Se o participante conhecer a palavra, deverá apertar a tecla P do teclado do computador. Se não a conhecer, deve apertar a tecla Q.
4. O participante é informado que fará, primeiramente, uma rodada de treino. Nesta rodada, ele deverá fazer o julgamento de oito palavras.
5. Após a etapa de treino, o participante realiza a mesma tarefa com mais dados. Nesta etapa, ele será apresentado a 30 itens de modo aleatório: 20 palavras e 10 itens distratores

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br



Continuação do Parecer: 6.015.577

(pseudopalavras e não-palavras).

Como Hipótese, a autora acredita que o tempo de reação na tarefa de decisão lexical será menor quando o prime for relacionado ao alvo, ou seja, quando o prime for uma palavra-base relacionada ao truncamento alvo. Acreditamos ainda, de acordo com os resultados de Heineck (2018) que fatores como a tonicidade e o número de sílabas da palavra-base podem ter influência na formação de truncamentos..

Tamanho da amostra: 150 participantes.

Critério de Inclusão:

Poderão participar do experimento falantes nativos do português brasileiros, destros e maiores de 18 anos. É importante que o participante seja destro, de modo a termos o controle de que todas as respostas afirmativas (clique na letra P) sejam dadas com a mão dominante do participante e que esse não constitua em um fator atrapalhador.

Critério de Exclusão:

Serão excluídos os participantes que residiram por períodos maiores que 5 anos em outro país.

Orçamento:

A pesquisa prevê um orçamento de R\$ 200,00, com financiamento próprio.

Cronograma:

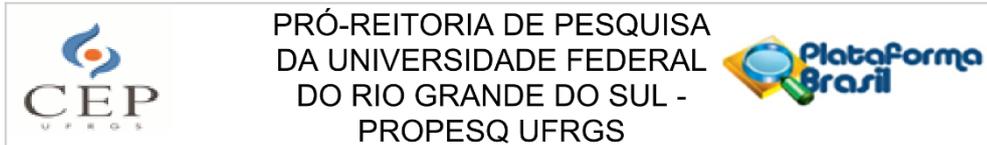
Informações da PB e Projeto Detalhado: Etapa de "Coleta e análise de dados" prevista para iniciar 25/04/2023.

Objetivo da Pesquisa:

Objetivo primário:

O objetivo geral da pesquisa é investigar e descrever o funcionamento do processo de truncamento em português brasileiro e promover uma discussão teórico-analítica das formas de base e das restrições fonológicas e morfológicas que dão conta do processo, de modo a contribuir para a discussão a respeito do funcionamento deste processo e para o seu entendimento do ponto de vista do seu enquadramento na gramática do português e das línguas naturais de um modo

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br



Continuação do Parecer: 6.015.577

geral.

Objetivos secundários:

Como objetivos específicos, podemos mencionar:

- (i) Compor um corpus de dados atuais de truncamentos que possa contribuir para os corpora de dados deste processo já existentes;
- (ii) Investigar de que modo se dá o processamento morfológico do truncamento por falantes do português brasileiro, através de um experimento virtual de decisão lexical com priming;
- (iii) Discutir o papel de fatores fonológicos como a tonicidade e o número de sílabas para a formação de truncamentos;
- (iv) Investigar o processamento morfológico de truncamentos formados a partir de palavras compostas por duas raízes eruditas (psicologia).

Avaliação dos Riscos e Benefícios:

Riscos:

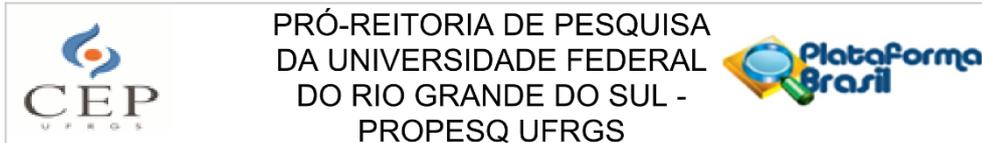
A participação na pesquisa não deve ocasionar nenhum dano moral aos participantes. No entanto, alguns participantes podem apresentar desconforto físico pelo tempo de exposição a tela do computador e a concentração exigida pelo experimento. Esse risco, caso se concretize, será de caráter passageiro, e não permanente. Procuramos minimizar este risco solicitando que o participante realize o experimento em um computador, ao invés de um aparelho eletrônico (como celular ou tablet, por exemplo). Desse modo, o participante pode sentar-se de forma mais confortável e realizar o experimento em uma tela maior. A tarefa também foi desenvolvida de forma otimizada quanto ao seu tempo de execução, permitindo que o participante a realize em cerca de 5 minutos. O participante, tomando conhecimento dos objetivos da pesquisa através do TCLE, poderá decidir por não participar da pesquisa caso julgá-la potencialmente danosa intelectual ou moralmente. O participante também tem a opção de encerrar sua participação no experimento a qualquer momento.

Benefícios

O participante poderá ser beneficiado uma vez que sua participação no experimento possivelmente gerará reflexões pessoais de ordem metalinguística acerca de seu próprio idioma. Também, será beneficiado indiretamente uma vez que o avanço das discussões analítico- teóricas na área dos estudos da linguagem pode fornecer benefício pedagógico e científico à sociedade, na

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br

Página 04 de 10



Continuação do Parecer: 6.015.577

qual o participante está incluso, uma vez que os resultados do experimento serão divulgados e publicados em eventos de ordem científica. Não existem, entretanto, outros benefícios diretos ao participante da pesquisa.

Comentários e Considerações sobre a Pesquisa:

Vide campo 'Conclusões ou Pendências e Lista de Inadequações'.

Considerações sobre os Termos de apresentação obrigatória:

1ª. versão:

TCLE apresentado.

PROJETO DETALHADO apresentado.

INFORMAÇÕES BÁSICAS DA PB. apresentadas.

2ª. versão

TCLE apresentado.

PROJETO DETALHADO apresentado.

INFORMAÇÕES BÁSICAS DA PB. Documento apresentado.

Carta-resposta. Apresentada.

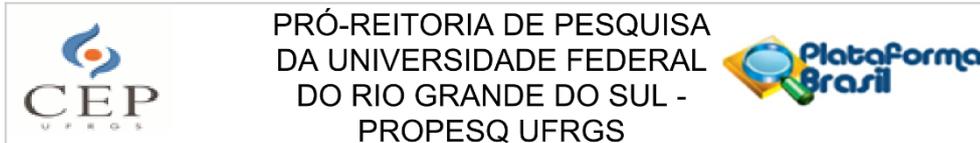
Conclusões ou Pendências e Lista de Inadequações:

Trata-se de uma resposta ao parecer consubstanciado CEP n.º 5.974.934, datado de 30/04/2023:

1) PLATAFORMA BRASIL (PB) E PROJETO DETALHADO:

1.a) O cronograma do estudo não está adequado, pois informa que ele já teria iniciado. Ressalta-se que a conduta do Sistema CEP/Conep tem sido de não emitir parecer em pesquisas concluídas ou em andamento. Tal decisão baseia-se no fato de o parecer ético não ser algo meramente burocrático, mas uma contribuição para a adequação do projeto de pesquisa às normas éticas vigentes, protegendo, assim, os interesses dos participantes e, conseqüentemente, de todos os envolvidos no processo: pesquisador, instituição, CEP e o próprio Sistema CEP/Conep. Solicita-se esclarecimentos e, se necessário, adequação do cronograma em relação à data de início do estudo (Resolução CNS n.º 466, de 2012, item XI.2.a). Solicita-se inserir, no projeto detalhado e na Plataforma Brasil, o compromisso explícito do pesquisador de que a pesquisa somente será iniciada a partir da aprovação pelo Sistema CEP/Conep (Norma Operacional CNS n.º 001, de 2013, item 3.4.1.9).

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br



PRÓ-REITORIA DE PESQUISA
DA UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL -
PROPESQ UFRGS

Continuação do Parecer: 6.015.577

RESPOSTA: O cronograma foi alterado de modo a iniciar somente após a aprovação pelo CEP. Inseriu-se um compromisso explícito de que a pesquisa será iniciada apenas a partir da aprovação pelo Sistema CEP/Conep. As alterações estão nas páginas 17 e 18 do projeto – e também foi realizada na Plataforma Brasil.

ANÁLISE: PENDÊNCIA ATENDIDA.

1.b) No projeto e na Plataforma Brasil, menciona-se que o convite será enviado por meio de redes sociais Twitter, Instagram e Facebook e por meio de contatos pessoais. Qualquer convite individual, enviado por e-mail ou contato pessoal, só poderá ter um único remetente e destinatário, ou ser enviado na forma de lista oculta. Dessa forma, solicitam-se esclarecimentos acerca do convite pessoal e, caso necessário, adequação (Carta Circular n.º1/2021-CONEP/SECNS/MS, item 2.1.1).

RESPOSTA: Foi esclarecido no projeto que os convites pessoais serão enviados por e-mail na forma de lista oculta. A alteração foi realizada na página 15 do projeto. Na PB a informação não foi mencionada.

ANÁLISE: PENDÊNCIA ATENDIDA.

1.c) Solicita-se que sejam inseridos os links no Projeto Detalhada e na PB (se for mais de um endereço eletrônico) para acesso ao questionário e ao TCLE que serão disponibilizados aos participantes.

RESPOSTA: Os links para os experimentos e para o TCLE que serão disponibilizados aos participantes foram adicionados ao projeto na página 13. Todos os links, incluindo o que direciona ao TCLE, foi incluído.

ANÁLISE: PENDÊNCIA ATENDIDA.

1d) No projeto e na PB há a menção que o participante deve ser destro, no entanto, essa informação não é esclarecida. Solicitam-se esclarecimentos a respeito desse critério de inclusão.

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br

RESPOSTA: Foi esclarecido o critério de inclusão relativo ao participante ser destro na página 15. “É importante que o participante seja destro, de modo a termos o controle de que todas as respostas afirmativas (clique na letra P) sejam dadas com a mão dominante do participante e que esse não constitua em um fator atrapalhador”.

ANÁLISE: PENDÊNCIA ATENDIDA.

2) TCLE E DADOS COLETADOS:

2.a) Quando a coleta de dados ocorrer em ambiente virtual (com uso de programas para coleta ou registro de dados, e-mail, entre outros), na modalidade de consentimento (Registro ou TCLE), o pesquisador deve enfatizar a importância do participante de pesquisa guardar em seus arquivos uma cópia (Carta Circular n.º 1/2021-CONEP/SECNS/MS, item 2.2).

RESPOSTA: Foi acrescentado ao TCLE o pedido de que o participante guarde em seus arquivos uma cópia do Termo. Alteração pode ser visualizada na página 23 do Projeto Detalhado.

ANÁLISE: PENDÊNCIA ATENDIDA

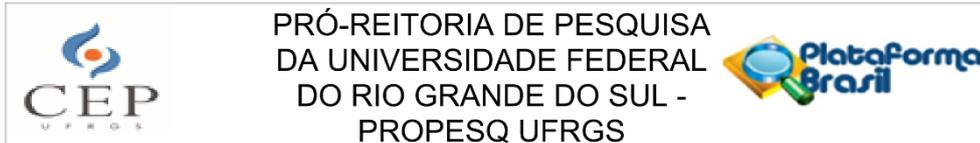
2.b) É de responsabilidade do pesquisador o armazenamento adequado dos dados coletados, bem como os procedimentos para assegurar o sigilo e a confidencialidade das informações do participante da pesquisa. Sendo assim, solicitam-se esclarecimentos acerca do armazenamento dos dados coletados no estudo (Carta Circular n.º 1/2021-CONEP/SECNS/MS, item 3.1).

RESPOSTA: O modo de armazenamento dos dados coletados foi esclarecido na página 16. “Uma vez concluída a coleta de dados, será feito o download dos dados gerados para o computador da pesquisadora, e serão excluídos da plataforma virtual no qual foram coletados.”

ANÁLISE: PENDÊNCIA ATENDIDA.

2.c) Inserir as informações completas e corretas do Comitê de Ética no TCLE, colocando o telefone correto ao final do documento: "O projeto foi avaliado pelo CEP-UFRGS, órgão colegiado,

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br



PRÓ-REITORIA DE PESQUISA
DA UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL -
PROPESQ UFRGS

Continuação do Parecer: 6.015.577

de caráter consultivo, deliberativo e educativo, cuja finalidade é avaliar – emitir parecer e acompanhar os projetos de pesquisa envolvendo seres humanos, em seus aspectos éticos e metodológicos, realizados no âmbito da instituição. O CEP UFRGS está localizado na Av. Paulo Gama, 110, Sala 311, Prédio Anexo I da Reitoria – Campus Centro, Porto Alegre/RS - CEP: 90040-060. Fone: +55 51 3308 3787 E-mail: etica@propesq.ufrgs.br Horário de Funcionamento: de segunda a sexta, das 08:00 às 12:00 e das 13:30 às 17:30h").

RESPOSTA: As informações completas e corretas do Comitê de Ética foram acrescentadas ao TCLE e podem ser vistas na página 23 do Projeto Detalhado.

ANÁLISE: PENDÊNCIA ATENDIDA.

2.d) Solicita-se inserir no TCLE a explicitação acerca do direito de buscar indenização diante de eventuais danos decorrentes da pesquisa (Resolução CNS n.º 466, de 2012, item IV.3.h).

RESPOSTA: Foi acrescentada ao TCLE a informação acerca da possibilidade de o participante buscar indenização (página 23 do Projeto Detalhado). “A assinatura deste Termo não exclui a sua possibilidade de buscar indenização diante de eventuais danos decorrentes de participação na pesquisa, como preconiza a Resolução 466/12.”

ANÁLISE: PENDÊNCIA ATENDIDA.

2.e) Solicita-se que conste, no consentimento (Registro ou TCLE), que nos casos em que não for possível a identificação do participante pelo questionário, o pesquisador deverá esclarecer a impossibilidade de exclusão dos dados da pesquisa durante o processo de registro/consentimento (Carta Circular n.º 1/2021-CONEP/SECNS/MS, item 4.3).

RESPOSTA: Ficou esclarecido no TCLE que, caso o questionário do participante não possa ser identificado, há a impossibilidade de exclusão dos dados da pesquisa (página 23 do TCLE).

ANÁLISE: PENDÊNCIA ATENDIDA.

Todas as pendências foram atendidas, não sendo observados óbices éticos nos documentos do

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br

Continuação do Parecer: 6.015.577

estudo.

Diante do exposto, o Comitê de Ética em Pesquisa – CEP-UFRGS, de acordo com as atribuições definidas na Resolução CNS n.º 510, de 2016, na Resolução CNS n.º 466, de 2012, e na Norma Operacional n.º 001, de 2013, do CNS, manifesta-se pela aprovação do protocolo de pesquisa proposto.

Reitera-se aos pesquisadores a necessidade de elaborar e apresentar os relatórios parciais e final da pesquisa, como preconiza a Resolução CNS/MS nº 466/2012, Capítulo XI, Item XI.2: "d.

Considerações Finais a critério do CEP:

Aprovado.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2091297.pdf	08/04/2023 17:28:47		Aceito
Outros	CartaRespostaCEP_DeboraHeineck.pdf	08/04/2023 17:28:15	DEBORA HEINECK	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_ComiteEtica_DeboraHeineck_atualizado_CEP.pdf	08/04/2023 17:27:11	DEBORA HEINECK	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_atualizadoCEP.pdf	08/04/2023 17:26:40	DEBORA HEINECK	Aceito
Folha de Rosto	FolhaDeRostoDeboraHeineck.pdf	26/02/2023 16:11:01	DEBORA HEINECK	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br



PRÓ-REITORIA DE PESQUISA
DA UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL -
PROPESQ UFRGS



Continuação do Parecer: 6.015.577

PORTO ALEGRE, 22 de Abril de 2023

Assinado por:
Patrícia Daniela Melchiors Angst
(Coordenador(a))

Endereço: Av. Paulo Gama, 110 - Sala 311 do Prédio Anexo 1 da Reitoria - Campus Centro
Bairro: Farroupilha **CEP:** 90.040-060
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3308-3787 **E-mail:** etica@propesq.ufrgs.br

Página 10 de 10

10.2 CONVITE PARA PARTICIPAÇÃO EM PESQUISA

**Pesquisa sobre
palavras do português**

REQUISITOS

- Ser brasileiro e falante de português brasileiro
- Ser destro
- Ter acima de 18 anos
- Ter acesso a um computador ou laptop

SOBRE A PESQUISA

O experimento dura em torno de 5 minutos para ser respondido e sua única tarefa é julgar o mais rápido possível se a palavra apresentada pode ou não pode ser usada no português. Durante o experimento, você não pode fazer nenhuma atividade paralela, então somente inicie quando você tiver 5 minutos disponíveis, ok? Todas as demais instruções necessárias serão dadas quando você clicar no link.

Clique no link na descrição

Muito obrigada!

10.3 TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DO EXPERIMENTO DE PROCESSAMENTO MORFOLÓGICO

Programa de Pós-Graduação em Letras UFRGS

Estudos da Linguagem: Fonologia e Morfologia

Você está sendo convidado(a) a participar da pesquisa intitulada Investigação experimental sobre o fenômeno do truncamento no português brasileiro, sob a responsabilidade do pesquisador e orientador Dr. Luiz Carlos da Silva Schwindt e condução da pesquisadora Débora Heineck, doutoranda do curso de pós-graduação em Letras. Para participar, você precisa ser destro, brasileiro(a), falante de português brasileiro, ter acima de 18 anos e também precisa ter acesso a um computador.

O objetivo deste trabalho é investigar o funcionamento e o processamento de um fenômeno de formação de palavras do português. O projeto foi avaliado pelo CEP-UFRGS, órgão colegiado, de caráter consultivo, deliberativo e educativo, cuja finalidade é avaliar – emitir parecer e acompanhar os projetos de pesquisa envolvendo seres humanos, em seus aspectos éticos e

metodológicos, realizados no âmbito da instituição. O CEP UFRGS está localizado na Av. Paulo Gama, 110, Sala 311, Prédio Anexo I da Reitoria – Campus Centro, Porto Alegre/RS - CEP: 90040-060. Fone: +55 51 3308 3787 E-mail: etica@propeq.ufrgs.br Horário de Funcionamento: de segunda a sexta, das 08:00 às 12:00 e das 13:30 às 17:30h.

Para realizar o estudo, será necessário que você se disponibilize a participar de uma tarefa de leitura silenciosa de palavras projetadas na tela de um computador e no julgamento sobre esta ser ou não uma palavra da língua portuguesa. Cabe dizer que não se trata de um teste de leitura; serão medidas apenas informações relativas ao tempo e ao índice de acertos da decisão lexical.

Os dados capturados passarão a compor o *corpus* de análise deste estudo, que engloba uma tese de doutorado em andamento e artigos científicos provenientes da tese. Serão solicitadas informações como sua idade e sua escolaridade; entretanto, elas não possuem o objetivo de identificar o participante. O presente experimento possui caráter anônimo, portanto, todas as informações coletadas nesta pesquisa são confidenciais. Trataremos todas as informações sem que haja identificação de particularidades de cada participante. Caso sinta algum tipo de desconforto, é possível parar a sua participação a qualquer momento e fazer uma pausa para retornar ao questionário. A assinatura deste Termo não exclui a sua possibilidade de buscar indenização diante de eventuais danos decorrentes de participação na pesquisa, como preconiza a Resolução 466/12. Também, informamos que os dados aqui coletados serão guardados pelos pesquisadores responsáveis por, no mínimo, 5 anos.

A pesquisa é feita de forma voluntária. Em virtude das informações coletadas serem utilizadas unicamente com fins científicos, garantimos o total sigilo e confidencialidade, através da concordância com este termo. Ademais, você terá direito e a liberdade de negar-se a participar desta pesquisa total ou parcialmente ou dela retirar-se a qualquer momento, sem que isto lhe traga qualquer prejuízo, de acordo com a Resolução CNS 510/2016 e complementares. Você também possui o direito de acessar os resultados desta pesquisa, que ficarão disponíveis com os responsáveis da pesquisa, conforme garante a Resolução CNS n.º 510, de 2016, Artigo 17, Inciso VI. Visto a confidencialidade e anonimidade dos dados, após a realização do questionário, não é possível identificar o questionário do participante e, portanto, também não é possível excluir os dados da pesquisa do participante. Uma vez que este experimento será realizado por meio de um questionário virtual, é importante que você guarde em seus arquivos uma cópia deste documento eletrônico de anuência.

Para qualquer esclarecimento no decorrer da sua participação, favor contatar o pesquisador responsável (schwindt@ufrgs.br) ou a pesquisadora assistente (heineck.debora@gmail.com) por e-mail ou no endereço Avenida Bento Gonçalves, 9500 – Agronomia – Porto Alegre/RS – Instituto

de Letras, sala 212 ou o Comitê de Ética em Pesquisa da universidade pelo telefone (51) 33083738. Desde já agradecemos!

Ao concordar em prosseguir com a pesquisa, você declara que, após ter sido esclarecido(a) pelo(a) pesquisador(a), lido o presente termo, e entendido tudo o que foi explicado, concorda em participar da Pesquisa intitulada Investigação experimental sobre o fenômeno do truncamento no português brasileiro.

Pesquisador responsável

Luiz Carlos da Silva Schwindt

schwindt@ufrgs.br

Pesquisadora assistente

Débora Heineck

heineck.debora@gmail.com

Trecho apresentado na plataforma Psytoolkit:

Esse estudo tem como objetivo investigar o funcionamento de um processo da língua portuguesa conhecido no meio linguístico por truncamento. Você deverá fazer uma leitura silenciosa das palavras que aparecem em sua tela e avaliar, o mais rápido possível, se a palavra apresentada pode ou não pode ser usada em português. Todas as instruções necessárias serão dadas ao longo do experimento.

O Termo de Consentimento Livre e Esclarecido está presente no seguinte link: <https://1drv.ms/b/s!AkOYizCVjBaKgvMm9DB8Ih-W7MLm1g?e=4LVG5v>

Ao concordar em prosseguir com a pesquisa, você declara que, após ter lido o presente termo e entendido tudo o que foi explicado, concorda em participar da pesquisa intitulada Processos de encurtamento vocabular em português brasileiro.

Eu compreendo as condições desta pesquisa.

10.4 DADOS UTILIZADOS NO EXPERIMENTO

- Palavra-base trissílaba oxítone → truncamento dissílabo oxítone

<i>Prime</i>	<i>Alvo</i>
--------------	-------------

condição	condí
visual	visú
guarujá	Guarú
carnaval	Carná

- Palavra-base trissílaba oxítona → truncamento dissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
juvenil	júve
futebol	fúte
chantilly	chânti
coração	córas
japonês	jápa

- Palavra-base trissílaba oxítona → truncamento trissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
capitão	Capíta
labrador	lábra
Futebol	futíba
comissão	comíssa
condição	condíça

- Palavra-base trissílaba paroxítona → truncamento dissílabo oxítono

<i>Prime</i>	<i>Alvo</i>
beleza	belê
mocreia	mocré
fissura	fissú
colégio	colé

- Palavra-base trissílaba paroxítona → truncamento dissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
gorjeta	gúrja
presente	préza
preguiça	prêgui
piscina	Písça
bobeira	bóbis

- Palavra-base polissílaba oxítona → truncamento dissílabo oxítono

<i>Prime</i>	<i>Alvo</i>
delineador	delí
disposição	dispô
exposição	expô
depilação	depí

- Palavra-base polissílaba oxítona → truncamento trissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
consumação	consúma
satisfação	satísfa
vestibular	vestíba
maracanã	maráca
ventilador	ventíla

- Palavra-base polissílaba oxítona → truncamento dissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
dissertação	díssers
qualificação	quáli

- Palavra-base polissílaba paroxítona → truncamento dissílabo oxítono

<i>Prime</i>	<i>Alvo</i>
jabuticaba	jabú
cahoeira	cachú
troglodita	troglô
combinado	combí
profissão	profí
mordomia	mordô

- Palavra-base polissílaba paroxítona → truncamento trissílabo oxítono

<i>Prime</i>	<i>Alvo</i>
possibilidade	possíbí
internet	internê
silicone	silicô
anabolizante	anabol

- Palavra-base polissílaba paroxítona → truncamento trissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
adolescente	adôle
deficiente	defíça
telefone	teléfa
adrenalina	adrêna
estrangeiro	estrânja
categoria	catéga

- Palavra-base polissílaba paroxítona → truncamento dissílabo paroxítono

<i>Prime</i>	<i>Alvo</i>
--------------	-------------

pervertido	pêrva
pindaíba	pínda
publicidade	públi
confiança	cônfa
barcelona	bárça

- Palavra-base composta 2 palavras morfológicas → truncamento

Prime	Alvo
hematologista	hemáto
poliomelite	pólio
pedagogia	pedágo
paleontologia	páleo
espeleologia	espéleo

- Não palavras

Prime	Alvo
swixte	swíx
chewp	êwp
ghyle	ghýl
kaphte	phté
dartch	dártc
gooklor	góok
ihnund	ihnún
kaustr	kaúst
dochew	chêw
werhor	wérh
scawnt	scáw
dumphs	úmphs
gnilt	gníl
thwolb	thwô
cwupte	cwúp
eccenblim	eccênb
schemicaccers	schemíc
emmetr	mêtr
inquat	quát
sarihy	ríhy
thicksiltles	thícks
geyflorin	florrín
huggle	húgg
czatativ	czát
oryoplec	oryôp
elytomet	tomét
ngwatibi	ngwát
aptordle	tórdle
jockirid	jóck
unjonitt	jonnít

10.5 MODELO DE *SCRIPT* UTILIZADO NA TAREFA DE DECISÃO LEXICAL COM *PRIMING*

O *script* apresenta os itens da lista 1 e está dividido em duas colunas.

options			readkey @5 5000
fullscreen			clear 2 3
background color black			if STATUS == CORRECT
			delay 1000
bitmaps			clear 5
instruction1			fi
instruction2			if STATUS != CORRECT
instruction3			delay 1000
instruction4			clear 5
instruction5			fi
fixpoint			save BLOCKNAME TABLEROW @1 @2
			@3 @4 @5 RT TT STATUS
fonts			
arial 40			table words
			#BASE TRISSILABA OXITONA
table training			TRUNCAMENTO DISSILABO OXITONO
"MONOGRAFIA" "MONÔ" "related"			"CONDIÇÃO" "CONDÍ" "related"
"training" 1			"BaseTriOxi" "TrunDiOxi" 1
"OPPILUME" "ÓPPIL" "related"			"VISUAL" "VISÚ" "related"
"training" 2			"BaseTriOxi" "TrunDiOxi" 1
"ANABOLIZANTE" "ANTÍFA"			#BASE TRISSILABA OXITONA
"unrelated" "training" 1			TRUNCAMENTO DISSILABO
"OCCLON" "CLÓN" "related"			PAROXITONO
"training" 2			"JUVENIL" "JÚVE" "related"
"QUERIDA" "QUÊRIS" "related"			"BaseTriOxi" "TrunDiParo" 1
"training" 1			#BASE TRISSILABA OXITONA
"ABSTANS" "ÁBST" "related"			TRUNCAMENTO TRISSILABO
"training" 2			PAROXITONO
"ESTAGIÁRIO" "PÚBLI" "unrelated"			"PEDAGOGIA" "CAPÍTA" "unrelated"
"training" 1			"BaseTriOxi" "TrunTriParo" 1
"SMOOMS" "SMÓO" "related"			"LABRADOR" "LÁBRA" "related"
"training" 2			"BaseTriOxi" "TrunTriParo" 1
			#BASE TRISSILABA PAROXITONA
task training			TRUNCAMENTO DISSILABO OXITONO
table training			"BELEZA" "BELÊ" "related"
keys p q			"BaseTriParo" "TrunDiOxi" 1
show bitmap fixpoint			#BASE TRISSILABA PAROXITONA
delay 500			TRUNCAMENTO DISSILABO
clear -1			PAROXITONO
show text @1 0 -25 255 255 255			"GORJETA" "GÚRJA" "related"
delay 50			"BaseTriParo" "TrunDiParo" 1
clear -1			"DEFICIENTE" "PRÉZA" "unrelated"
show text @2 0 -25 255 255 255			"BaseTriParo" "TrunDiParo" 1

```

#BASE POLISSILABA OXITONA
TRUNCAMENTO DISSILABO OXITONO
"DELINEADOR" "DELÍ" "related"
"BasePoliOxi" "TrunDiOxi" 1
#BASE POLISSILABA OXITONA
TRUNCAMENTO TRISSILABO
PAROXITONO
"CONSUMAÇÃO" "CONSÚMA" "related"
"BasePoliOxi" "TrunTriParo" 1
#BASE POLISSILABA OXITONA
TRUNCAMENTO DISSILABO
PAROXITONO
"DISSERTAÇÃO" "DÍSSERS" "related"
"BasePoliOxi" "TrunDiParo" 1
#BASE POLISSILABA PAROXITONA
TRUNCAMENTO DISSILABO OXITONO
"PRESENTE" "CACHÚ" "unrelated"
"BasePoliParo" "TrunDiOxi" 1
"PROFISSÃO" "PROFÍ" "related"
"BasePoliParo" "TrunDiOxi" 1
"MORDOMIA" "MORDÔ" "related"
"BasePoliParo" "TrunDiOxi" 1
#BASE POLISSILABA PAROXITONA
TRUNCAMENTO TRISSILABO OXITONO
"SILICONE" "SILICÔ" "related"
"BasePoliParo" "TrunTriOxi" 1
#BASE POLISSILABA PAROXITONA
TRUNCAMENTO TRISSILABO
PAROXITONO
"CACHOEIRA" "DEFÍÇA" "unrelated"
"BasePoliParo" "TrunTriParo" 1
"ESTRANGEIRO" "ESTRÂNJA" "related"
"BasePoliParo" "TrunTriParo" 1
#BASE POLISSILABA PAROXITONA
TRUNCAMENTO DISSILABO
PAROXITONO
"PINDÁIBA" "PÍNDA" "related"
"BasePoliParo" "TrunDiParo" 1
"BARCELONA" "BÁRÇA" "related"
"BasePoliParo" "TrunDiParo" 1
#BASE 2 PALAVRAS MORFOLOGICAS
TRUNCAMENTO
"CAPITÃO" "PEDÁGO" "unrelated"
"Base2palavras" "Trun2palavras" 1
#NONWORDS
"SWIXTE" "SWÍX" "related" "nonword"
"nontrunc" 2
"CHEWP" "ÊWP" "related" "nonword"
"nontrunc" 2
"GHYLC" "GHÝL" "related" "nonword"
"nontrunc" 2
"KAPHTE" "PHTÉ" "related" "nonword"
"nontrunc" 2
"DARTCH" "DÁRTC" "related"
"nonword" "nontrunc" 2
"GOOKLOR" "GÓOK" "related"
"nonword" "nontrunc" 2
"IHNUND" "IHNÚN" "related"
"nonword" "nontrunc" 2
"KAUSTR" "KAÚST" "related"
"nonword" "nontrunc" 2
"DOCHEW" "CHÊW" "related"
"nonword" "nontrunc" 2
"WERHOR" "WÉRH" "related"
"nonword" "nontrunc" 2
task words
table words
keys p q
show bitmap fixpoint
delay 500
clear -1
delay 500
show text @1 0 -25 255 255 255
delay 50
clear -1
draw off
draw on
show text @2 0 -25 255 255 255
readkey @6 5000
clear 2 3
if STATUS == CORRECT
delay 500
clear 5
fi
if STATUS != CORRECT
delay 500
clear 5
fi
save BLOCKNAME TABLEROW @1 @2
@3 @4 @5 @6 RT TT STATUS
block training
message instruction1

```

message instruction2
 message instruction3
 message instruction4
 tasklist
 training 6 all_before_repeat
 end

block words
 message instruction5
 tasklist
 words 30 all_before_repeat
 end

10.6 ITENS COM ZERO OCORRÊNCIAS NO WEB/DIALECTS EM NOSSO CORPUS DE DADOS DE TRUNCAMENTOS

Truncamento	Palavra-base	Ocorrências em <i>tokens</i>
milha	milhão	0
mulha	mulher	0
condi	condição	0
cafu	cafune	0
trava	travesti	0
juve	juvenil	0
coras	coração	0
condiça	condição	0
sacrista	sacristão	0
labra	labrador	0
cafeta	cafetão	0
comissa	comissão	0
juvena	juvenil	0
mocre	mocreia	0
cole	colégio	0
fissu	fissura	0
justa	justiça	0
gurja	gorjeta	0
sarja	sargento	0
pisça	piscina	0
bermas	bermuda	0
berma	bermuda	0
bira	birita	0
preza	presente	0
catcha	cachaça	0
mendi	mendigo	0
falsifi	falsificação	0
deli	delineador	0
intima	intimação	0
ventila	ventilador	0
desperta	despertador	0
dissers	dissertação	0
retro	retropesctiva	0
quiti	quitinete	0
provi	(Morro da) Providência	0
alo	alojamento	0
razu	razoável	0
jabu	jabuticaba	0
metal	metaleiro	0

mordo	mordomia	0
troglo	troglodita	0
jabo	jaboticaba	0
mare	maresia	0
sante	Santa Tereza	0
combi	combinado	0
possibi	possibilidade	0
aspira	aspirante	0
catega	categoria	0
sanducha	sanduíche	0
carpina	carpinteiro	0
comissa	comissário	0
cortesa	cortesia	0
maneca	manequim	0
proletra	proletariado	0
tefafa	telefone	0
cafita	cafetina	0
defiça	deficiente	0
farroupis	farroupilha	0
pinda	pindaíba	0
cafaja	cafajeste	0
malcra	malcriado	0
estroguis	estregonofe	0
pedago	pedagogia	0
espeleo	espeleologia	0
over	overnight	0
portoale	Porto Alegre	0
flori	floricultura	0
tricampa	tricampeão	0
cupu	cupuaçu	0
dire	diretor	0
peri	periferia	0
delis	delícia	0
vontis	vontade	0

10.7 LISTA DE DADOS QUE COMPÕEM NOSSO CORPUS DE DADOS DE TRUNCAMENTO

Truncamento	Palavra-base
burga	burguês
china	chinês
milha	milhão
mulha	mulher
tusta	tostão
playba	playboy
brasa	brasil
condi	condicionador
condi	condição
visu	visual
cafu	cafuné
guaru	guarujá

carna	carnaval
japa	japonês
trava	travesti
confa	confusão
bisa	bisavo
chanti	chantilly
juve	juventus
juve	juventude
juve	juvenil
carna	carnaval
fute	futebol
coras	coração
condiça	condição
capita	capitão
sacrista	sacristão
sapata	sapatão
labra	labrador
boteco	botequim
cafeta	cafetão
celula	celular
comissa	comissão
juvena	juvenil
natureba	natural
futiba	futebol
futibas	futebol
futebas	futebol
traveco	travesti
bele	beleza
mocre	mocreia
cole	colégio
fissu	fissura
pregui	preguiça
pregui	preguiça
justa	justiça
neura	neurose
flagra	flagrante
prima	primeira
ceva	cerveja
cerva	cerveja
gurja	gorjeta
sarja	sargento
giga	gigante
granfa	grã-fino
churras	churrasco
cunha	cunhado
freela	free-lancer
pisça	piscina
bermas	bermuda
berma	bermuda
bira	birita
preza	presente
demo	demônio

pirras	pirralho
tranqui	tranquilo
tranquis	tranquilo
bobis	bobeira
catcha	cachaça
mendi	mendigo
miguís	amigo
trabis	trabalho
expo	exposição
depi	depilação
jaba	jabaculê
dispo	disposição
falsifi	falsificação
deli	delineador
intima	intimação
maraca	maracanã
satisfa	satisfação
consuma	consumação
vestiba	vestibular
ventila	ventilador
desperta	despertador
ilustra	ilustração
dissers	dissertação
quali	qualificação
biju	bijuteria
refri	refrigerante
bici	bicicleta
bici	bicicleta
preju	prejuízo
retro	retrospectiva
falsi	falsificado
quiti	quitinete
provi	(Morro da) Providência
belo	Belo Horizonte
alo	alojamento
razu	razoável
jabu	jabuticaba
profi	profissional
kili	Kilimanjaro
metal	metaleiro
mordo	mordomia
troglo	troglodita
jabo	jabuticaba
mare	maresia
sante	Santa Tereza
cachu	cachoeira
combi	combinado
possibi	possibilidade
interne	internet
silico	silicone
anabol	anabolizante
motora	motorista

salafra	salafrário
responosa	responsabilidade
estranja	estrangeiro
madruga	madrugada
batera	baterista
aspira	aspirante
analfa	analfabeto
vagaba	vagabunda
catega	categoria
ibira	ibirapuera
bonsuça	Bom Sucesso
adrena	adrenalina
delega	delegado
retarda	retardado
comuna	comunista
sanducha	sanduíche
balzaca	balzaquiana
carpina	carpinteiro
comissa	comissário
cortesa	cortesia
delega	delegacia
maneca	manequim
motoca	motocicleta
natura	natureza
perifa	periferia
proletra	proletariado
reaçã	reacionário
telefa	telefone
tráfica	traficante
araça	araçatuba
califa	califórnia
cafita	cafetina
gasosa	gasolina
presida	presidente
acampis	acampamento
defiça	deficiente
adole	adolescente
burocra	burocracia
farroupis	farroupilha
coca	cocaína
pinda	pindamonhangaba
pinda	pindaíba
copa	copacabana
cafaja	cafajeste
cafa	cafajeste
confa	confiança
barça	barcelona
perva	perverso
malcra	malcriado
profe	professor
brinquis	brincadeira
chape	chapecoense

namor(s)	namorado
aniver	aniversário
niver	aniversário
bici	bicicleta
publi	publicidade
estroguis	estrogenofe
gastro	gastroenterologista
eco	ecografia
polio	poliomielite
eletro	eletrocardiograma
fono	fonoaudiólogo
cardio	cardiologista
odonto	odontologia
gineco	ginecologista
oftalmo	oftalmologista
ultra	ultrassonografia
pedago	pedagogia
biblio	biblioteconomia
paleo	paleontologia
psico	psicólogo
penta	pentacampeão
hetero	heterossexual
homo	homossexual
hemato	hematologista
tetra	tetracampeão
neuro	neurologista
esteto	estetoscópio
otorrino	otorrinolaringologista
lipo	lipoaspiração
super	supermercado
curta	curta-metragem
espeleo	espeleologia
foto	fotografia
hidro	hidromassagem
hidro	hidroginástica
infra	infraestrutura
inter	internacional
longa	longa-metragem
moto	motocicleta
over	overnight
porno	pornográfico
quimio	quimioterapia
radio	radioterapia
portoale	Porto Alegre
flori	floricultura
pós	pós-graduação
ex	ex-namorado
ex	ex-marido
bi	bicampeão
bi	bissexual
tri	tricampeão
vice	vice-presidente

vice	vice-governador
vice	vice-campeão
vice	vice-prefeito
tricampa	tricampeão
trans	transexual
cupu	cupuaçu
dire	diretor
facul	faculdade
loto	loteria
peri	periferia
promo	promoção
sussa	sossegado
coraçã	coração
recunha	reconhecimento
fofis	fofo
delis	delícia
mamis	mamãe
papis	papai
vontis	vontade

10.8 EXPERIMENTO ONLINE (HEINECK, 2018)

Exercício 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 em caixa alta (letra maiúscula: como em viSU) para indicar a sílaba mais forte (tônica) da palavra.

- viSU
- BURga
- conDI
- JApa
- diRE
- boTEco
- PLAYba
- saCRISa
- beLE
- CONfa
- fisSU
- CUnha
- exPO
- maRAca
- saTIsfa

- aGRI
- DIsser
- reFRI
- FLAgra
- proFI
- PÓS
- maDRUGa
- venTI
- moTOra
- PROfe
- aNIver
- FOno
- oFTALmo
- BI
- dePI

Exercício 2

A partir das palavras abaixo, sugira uma forma encurtada da palavra que lhe pareça soa bem na fala informal. Escreva em caixa alta (letra maiúscula) a sílaba ou vogal que você considera forte (tônica) em cada uma das palavras que você sugerir.

- macromolécula
- chinês
- tostão
- Carnaval
- flexibilidade
- travesti
- bisavó
- juvenil
- cafuné
- futebol
- preguiça
- neurose
- tranquilo
- particular
- disposição

- vestibular
- qualificação
- bijuteria
- prejuízo
- ventilador
- responsabilidade
- poligamia
- baterista
- confiança
- bicicleta
- odontologia
- hematologista
- pré-escola
- hermoграфия
- tricampeão

Exercício 3

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

- CEva
- comPUta
- LIpo
- peRifa
- BIra
- saTISfa
- NEUuro
- FREEla
- porCE
- teLEfa
- FOto
- MaRAca
- ofTALmo
- deLEga

- temPEra
- BÍblio
- porTUga
- MOto
- comuna
- Retro

10.9 EXPERIMENTO PRESENCIAL (HEINECK, 2018)

IMPORTANTE: Nas perguntas abaixo, você deve preencher as lacunas de acordo com as instruções de cada questão. Caso você não conheça alguma das palavras listadas, proponha ainda assim uma forma que você considera a ela assemelhada. Não há uma resposta certa ou errada, mas sim, a sua intuição sobre essas formas.

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 _____

zela _____

boteco _____

playba _____

sacrista _____

bele _____

confa _____

fissu _____

cunha _____

expo _____

marça _____

satisfa _____

instru _____

disser _____
refri _____
flagra _____
profi _____
pós _____
madruga _____
desperta _____
motora _____
profe _____
aniver _____
fono _____
oftalmo _____
bi _____
depi _____

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.

Vamos para o **carnaval** hoje!

É preciso ter **maturidade** para enfrentar os problemas.

O **travesti** almoçou no restaurante.

Minha **bisavó** veio da Itália.

O **juvenil** vai jogar amanhã.

Adoro um **cafuné**!

Jogamos **futebol** todo sábado.

Estou com uma **preguiça** hoje!

Essa **neurose** por limpeza não te faz bem.

Fique **tranquilo**! Tudo vai dar certo!

É preciso um bom **paladar** para degustar estes queijos.

Você está sempre com muita **disposição** pela manhã!

Ele vai prestar o **vestibular** neste fim de semana.

A **qualificação** da minha tese é na quinta-feira.

Essa **bijuteria** parece joia de verdade.

Eu saí no **prejuízo** naquele negócio.

Posso ligar o **ventilador**?

Cuidar de uma criança é muita **responsabilidade**.

A **monogamia** é pregada por várias religiões.

O **baterista** tocou muito bem hoje!

O João é de **confiança**.

Pedro gosta de andar de **bicicleta**.

Ela está estudando **odontologia**.

A amostra de sangue será analisada por um **hematologista**.

Meu filho está na **pré-escola**.

Os cientistas se detiveram à **cetografia**.

O meu time é **tricampeão** da Copa do Mundo.

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**.

Tem cada vez mais gente da **perifa** na Universidade.

Eles saíram para tomar umas **birra** ontem.

É muita **satisfa** estar aqui.

O **neuro** solicitou novos exames.

Contratei um **freela** para fazer a tradução para mim.

Vou abrir a **persi** para entrar um pouco de luz na sala.

Você pode atender o **telefa**, por favor?

Essa sua **foto** está muito bonita.

O jogo será no **Maraca**.

Ele precisa consultar um **oftalmo**.

O **delega** autorizou a prisão.

O **eleva** não estava funcionando no meu prédio.

Passei o dia estudando na **biblio**.

No Brasil, há muitas piadas com **potuga**.

Minha **moto** está estacionada nesta rua.

O **comuna** organizou uma manifestação.

A professora utilizou o **retro** em aula hoje.

10.10 LISTA DE PALAVRAS DE BUSCA REALIZADA NO TWITTER (HEINECK, 2021)

Exercício 1	Exercício 2	Exercício 3
bi aniver refri japa madruga bele maraca oftalmo profe	fute juve respon s odonto neura bisa carna traveco tranks	fotinho fotinha motinho motinha

flagra visu playba fono expo motora profi depi disser satisfa desperta	preju trava biju quali tosta juvena cafu bici pregui china confi hemato dispo futiba batera vestiba tricampa	
--	--	--