

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS**

EDUARDO ELISALDE TOLEDO

**PRODUTIVIDADE DA DERIVAÇÃO REGRESSIVA
NO PORTUGUÊS BRASILEIRO: UM ESTUDO SOBRE
A VOGAL FINAL**

Porto Alegre

2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
Estudos da Linguagem
Teoria e Análise Linguística
Linha de Pesquisa: Fonologia e Morfologia

**PRODUTIVIDADE DA DERIVAÇÃO REGRESSIVA NO
PORTUGUÊS BRASILEIRO: UM ESTUDO SOBRE A
VOGAL FINAL**

EDUARDO ELISALDE TOLEDO

Orientador: Prof. Dr. Luiz Carlos da Silva Schwindt

Tese de Doutorado em Fonologia e Morfologia,
apresentada como requisito parcial para obtenção do
título de Doutor pelo Programa de Pós-Graduação em
Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, junho de 2020.

CIP - Catalogação na Publicação

Toledo, Eduardo Elisalde
Produtividade da Derivação Regressiva no Português
Brasileiro: um estudo sobre a vogal final / Eduardo
Elisalde Toledo. -- 2020.
168 f.
Orientador: Luiz Carlos da Silva Schwindt.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2020.

1. Derivação regressiva. 2. Vogal temática. 3.
Frequência lexical. 4. Morfologia. 5. Morfofonologia.
I. Schwindt, Luiz Carlos da Silva, orient. II.
Título.

EDUARDO ELISALDE TOLEDO

**PRODUTIVIDADE DA DERIVAÇÃO REGRESSIVA NO
PORTUGUÊS BRASILEIRO: UM ESTUDO SOBRE A
VOGAL FINAL**

Esta tese foi analisada e julgada adequada para a obtenção do título de doutor em Teoria e Análise Linguística e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientador: Professor Doutor Luiz Carlos da Silva Schwindt

Aprovada em: 30/06/2020

BANCA EXAMINADORA

Prof. Dr. Leandro Zanetti Lara – UFRGS

Profa. Dra. Camila De Bona – IFSUL

Prof. Dr. Vitor Augusto Nóbrega – USP

Dedico esta tese à memória de meu pai, Elói Toledo dos Santos (1943-2020), que me ensinou a jogar xadrez e a amar os livros. Seu amor vai sempre me acompanhar.

AGRADECIMENTOS

A Danielle Nery Matheus, por me ensinar todos os dias a linguagem secreta que se esconde nas coisas simples da vida. Por ser a musa que semeia meus sonhos com um amor infinito e enriquece a minha imaginação com a poesia de seu encanto arrebatador.

A minha família, meus pais, Elói Toledo dos Santos e Elena Maria Elisalde, e minha irmã, Estela Elisalde Toledo, por terem me motivado, desde meus primeiros dias na escola, a sempre valorizar o conhecimento e a curiosidade por aprender coisas novas.

Aos meus sogros, Edilene Aparecida Nery Matheus e Gilberto Matheus, e a meu cunhado, Guilherme Nery Matheus, por me fazerem todos os dias sentir cada vez mais parte de uma família tão amorosa e feliz.

A meu orientador, professor Luiz Carlos Schwindt, por ter me oferecido a oportunidade de retornar ao doutorado. Por ter me convidado para participar de sua pesquisa quando eu era um aluno de graduação à deriva pelo curso de Letras, no meu primeiro ano, em 2004. Além disso, por ter sempre acreditado no meu potencial intelectual, me oferecendo a oportunidade de aprender a fazer ciência e por ser um modelo de profissional ético e curioso para mim.

À Universidade Federal do Rio Grande do Sul, por ser há mais de 15 anos o meu porto seguro, o local onde dei meus primeiros passos como pesquisador, onde tive minhas primeiras conquistas acadêmicas e onde amadureci e fixei minhas raízes intelectuais e políticas.

Ao Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul, por ter permitido meu retorno ao doutorado e ter oferecido um atendimento sempre atencioso.

Aos professores do Instituto de Letras por terem me oferecido um ensino tão qualificado e instigante, que me moldou como professor de língua portuguesa.

Aos alunos, professores e profissionais da Escola Municipal de Ensino Fundamental Aramy Silva, por valorizarem meu esforço diário como profissional da educação.

*Come writers and critics
Who prophesize with your pen
And keep your eyes wide
The chance won't come again
And don't speak too soon
For the wheel's still in spin
And there's no tellin' who that it's namin'
For the loser now will be later to win
For the times they are a-changin'
DYLAN, Bob*

RESUMO

Este estudo descreve a produtividade da derivação regressiva no português brasileiro (PB) sob a ótica da vogal final. Questionando a aparente aleatoriedade ou imprevisibilidade da seleção da vogal final (*o*, *a* ou *e*) nas palavras formadas por derivação regressiva (ex. *abandono* < *abandonar*; *combate* < *combater*; *fuga* < *fugir*), este estudo tem por objetivo descrever os fatores linguísticos envolvidos nos padrões de produtividade da derivação regressiva a partir da análise de três amostras do léxico: (i) léxico dicionarizado; (ii) léxico institucionalizado e (iii) léxico potencial. Para isso, esta pesquisa toma como base a noção de produtividade em dois âmbitos, *disponibilidade* e *rentabilidade* (CORBIN, 1987), e assume, como base para os procedimentos metodológicos, a noção de *armazenamento de exemplares no léxico* (PIERREHUMBERT, 2001). Em primeiro lugar, observou-se a disponibilidade no léxico dicionarizado do PB por meio dos Dicionários Eletrônicos *Aurélio* e *Houaiss*. Os dados coletados foram, então, codificados, a partir de fatores linguísticos. Posteriormente, verificou-se a frequência lexical no léxico institucionalizado (*Corpus Brasileiro*) e os padrões de produtividade no léxico potencial (experimento online). Os resultados do experimento online foram comparados aos padrões morfológicos e fonológicos encontrados no léxico dicionarizado e no léxico institucionalizado. Em termos gerais, há convergência na distribuição das três vogais finais nas duas amostras de léxico. Os resultados também apresentam convergências para os fatores linguísticos conjugação verbal, número de sílabas e *onset* da última sílaba, e divergências, em complexidade morfológica, vogal média tônica e modo de articulação do contexto precedente. O experimento online replica o padrão geral do PB de alta frequência de verbos da primeira conjugação e das vogal final *a*, além de também apresentar alta incidência da vogal *o*, o que pode ser reflexo da tensão entre o predomínio de palavras femininas terminadas em *a* no padrão geral dos substantivos no português brasileiro e a frequência superior de palavras masculinas terminadas em *o* nos nomes de verbais formados por derivação regressiva. Além dessas duas forças, padrão feminino geral (*a*) e padrão masculino para deverbais (*o*), parece ter um papel relevante na seleção da vogal final o padrão sonoro final TA/DA e TE/DE.

Palavras-chave: Derivação regressiva. Vogal temática. Frequência lexical. Morfologia. Morfofonologia.

ABSTRACT

This study describes the productivity of back-formation in Brazilian Portuguese (BP) from the perspective of the final vowel. Questioning the apparent randomness or unpredictability of the selection of the final vowel (*o*, *a* or *e*) in the words formed by back-formation (e.g. *abandono* ‘abandonment’ < *abandonar* ‘to abandon’; *combate* ‘combat’ < *combater* ‘to combat’; *fuga* ‘escape’ < *fugir* ‘to escape’), this study aims to describe the linguistic factors involved in the productivity patterns of back-formation from the analysis of three samples of the lexicon: (i) dictionarized lexicon; (ii) institutionalized lexicon and (iii) potential lexicon. For this, this research is based on the notion of productivity in two areas, availability and profitability (CORBIN, 1987), and assumes, as a basis for methodological procedures, the notion of stored exemplars in the lexicon (PIERREHUMBERT, 2001). First, the availability in the dictionarized lexicon of BP was observed through the Electronic Dictionaries *Aurélio* and *Houaiss*. The collected data were then coded based on linguistic factors. Subsequently, the lexical frequency in the institutionalized lexicon (*Corpus Brasileiro*) and the productivity patterns in the potential lexicon (online experiment) were verified. The results of the online experiment were compared to the morphological and phonological patterns found in the dictionarized lexicon and the institutionalized lexicon. In general terms, there is convergence in the distribution of the three final vowels in the two samples. The results also show convergences for the linguistic factors verbal conjugation, number of syllables and *onset* of the last syllable, and divergences, in morphological complexity, tonic mid vowel and manner of articulation of the preceding context. The online experiment replicates the BP general pattern of high frequency of verbs of the first conjugation and of the final vowel *a*, in addition to also presenting a high incidence of the vowel *o*, which may reflect the tension between the predominance of feminine nouns ending in *a* in the general pattern of nouns in Brazilian Portuguese and the higher frequency of masculine nouns ending in *o* in the verbal nouns formed by back-formation. In addition to these two forces, general feminine pattern (*a*) and masculine pattern for deverbal (*o*), the final sound pattern TA/DA and TE/DE seems to have a relevant role in the selection of the final vowel.

Keywords: Back-formation. Thematic Vowel. Lexical Frequency. Morphology. Morphophonology.

LISTA DE GRÁFICOS

Gráfico 1 - Distribuição temporal das vogais finais.....	67
Gráfico 2 - Comparação entre types e tokens	83
Gráfico 3 - Padrões de sequências finais para vogal o: types	96
Gráfico 4 - Padrões de sequências finais para vogal o: tokens	96
Gráfico 5 - Padrões de sequências finais para vogal a: types.....	98
Gráfico 6 - Padrões de sequências finais para vogal a: tokens.....	98
Gráfico 7 - Padrões de sequências finais para vogal e: tokens.....	99
Gráfico 8 - Padrões de sequências finais para vogal e: types.....	99
Gráfico 9 - Fator social: sexo dos informantes	101
Gráfico 10 - Fator social: faixa etária dos informantes	102
Gráfico 11 - Distribuição da vogal final na Questão 01: Fator sexo.....	102

LISTA DE FIGURAS

Figura 1 - Relação arbitrária entre vogal temática e vogal final	20
Figura 2 - Membros da Classe I.	49
Figura 3 - Membros da Classe II.....	49
Figura 4 - Membros da Classe III.....	50
Figura 5 - Dicionário Eletrônico Aurélio	52
Figura 6 - Dicionário Eletrônico Houaiss	53
Figura 7 - Mecanismo de busca do Dicionário Eletrônico Aurélio.....	53
Figura 8 - Mecanismo de busca do Dicionário Eletrônico Houaiss	54
Figura 9 - Mecanismo de busca do Corpus Brasileiro no site linguateca.pt	55
Figura 10 - Informações de frequência no Corpus Brasileiro	55

LISTA DE TABELAS

Tabela 1 - Distribuição da vogal final dos substantivos no Dicionário Eletrônico Aurélio	63
Tabela 2 - Distribuição das três vogais finais nos nomes deverbais	64
Tabela 3 - Distribuição das conjugações verbais nos nomes deverbais em geral	64
Tabela 4 - Palavras encontradas nos dois dicionários	65
Tabela 5 - Palavras sem regionalismos	65
Tabela 6 - Palavras sem formas com homofonia acidental	65
Tabela 7 - Palavras sem formas alternantes	66
Tabela 8 - Distribuição das vogais finais pela linha do tempo.....	66
Tabela 9 - Comparação entre critérios de organização dos dados	68
Tabela 10 - Distribuição de vogais finais nos casos de regionalismo	68
Tabela 11 - Regionalismos: frequência para vogal final o no Corpus Brasileiro	68
Tabela 12 - Regionalismos: frequência para vogal final a no Corpus Brasileiro.....	69
Tabela 13 - Regionalismos: frequência para vogal final e no Corpus Brasileiro.....	70
Tabela 14 - Formas com alternância: frequência para vogal final o no Corpus Brasileiro.....	70
Tabela 15 - Formas com alternância: frequência para vogal final a no Corpus Brasileiro	71
Tabela 16 - Formas com alternância: frequência para vogal final e no Corpus Brasileiro	71
Tabela 17 - Distribuição das formas alternantes nas três vogais finais: types	72
Tabela 18 - Distribuição das formas alternantes nas três vogais finais: tokens	73
Tabela 19 - Formas não listadas no Houaiss: frequência da vogal final o no Corpus Brasileiro.....	73
Tabela 20 - Formas não listadas no Houaiss: frequência da vogal final e no Corpus Brasileiro	74
Tabela 21 - Distribuição das palavras incluídas para as três vogais finais	74
Tabela 22 - Distribuição das vogais finais no dicionário e no Corpus Brasileiro.....	75
Tabela 23 - Fator Vogal Temática Verbal: distribuição das vogais finais	75
Tabela 24 - Fator Vogal Temática: distribuição das vogais finais	76
Tabela 25 - Fator Complexidade Morfológica: distribuição das vogais finais	77
Tabela 26 - Distribuição dos prefixos para a vogal final e.....	77
Tabela 27 - Distribuição dos prefixos para a vogal final o	77
Tabela 28 - Distribuição dos prefixos para a vogal final a.....	78
Tabela 29 - Fator Vogal Média Tônica: distribuição das vogais finais	79
Tabela 30 - Fator Gênero da Palavra: distribuição das vogais finais	79
Tabela 31 - Fator Número de Sílabas: distribuição das vogais finais	80
Tabela 32 - Fator Onset da Última Sílabas: distribuição das vogais finais	80

Tabela 33 - Fator Contexto Fonológico Precedente (modo de articulação): distribuição das vogais finais	81
Tabela 34 - Fator Contexto Fonológico Precedente (ponto de articulação): distribuição das vogais finais	82
Tabela 35 - Comparação entre tokens e types: distribuição das vogais finais	82
Tabela 36 - Fator vogal temática verbal: Corpus Brasileiro	83
Tabela 37 - Fator vogal temática: Corpus Brasileiro	84
Tabela 38 - Fator Complexidade Morfológica: Corpus Brasileiro	84
Tabela 39 - Fator Vogal Média Tônica: Corpus Brasileiro.....	85
Tabela 40 - Fator Gênero da Palavra: Corpus Brasileiro	85
Tabela 41 - Fator Número de Sílabas: Corpus Brasileiro	86
Tabela 42 - Fator Onset da Última Sílaba: Corpus Brasileiro.....	87
Tabela 43 - Fator Contexto Fonológico Precedente (modo de articulação): Corpus Brasileiro	87
Tabela 44 - Fator Contexto Fonológico Precedente (ponto de articulação): Corpus Brasileiro	88
Tabela 45 - Comparação entre types e tokens: fator vogal temática verbal.....	89
Tabela 46 - Comparação entre types e tokens: fator vogal temática.....	90
Tabela 47 - Comparação entre types e tokens: fator complexidade morfológica	90
Tabela 48 - Comparação entre types e tokens: fator vogal média tônica.....	91
Tabela 49 - Comparação entre types e tokens: fator gênero da palavra.....	91
Tabela 50 - Comparação entre types e tokens: fator gênero da palavra.....	92
Tabela 51 - Comparação entre types e tokens: fator onset da última sílaba	92
Tabela 52 - Comparação entre types e tokens: fator contexto fonológico precedente (modo de articulação).....	93
Tabela 53 - Comparação entre types e tokens: fator contexto fonológico precedente (ponto de articulação).....	93
Tabela 54 - Padrões de sequências finais para vogal o: comparação entre types e tokens	95
Tabela 55 - Padrões de sequências finais para vogal a: comparação entre types e tokens	97
Tabela 56 - Padrões de sequências finais para vogal e: comparação entre types e tokens	99
Tabela 57 - Respostas dos informantes à Questão 01	104
Tabela 58 - Respostas dos informantes à Questão 02	104
Tabela 59 - Respostas dos informantes à Questão 03	105
Tabela 60 - Respostas dos informantes à Questão 04	106
Tabela 61 - Respostas dos informantes à Questão 05	106
Tabela 62 - Respostas dos informantes à Questão 06	107
Tabela 63 - Cruzamento de respostas dos informantes: forma derivada x forma infinitiva	109
Tabela 64 - Cruzamento das respostas dos informantes: formas derivadas x formas derivadas	110

LISTA DE QUADROS

Quadro 1 - Fenômenos linguísticos associados à frequência das palavras	40
Quadro 2- Classes formais do português.....	47
Quadro 3 - Comparação entre types e tokens: convergências e divergências.....	94
Quadro 4 - Comparação entre tokens e types: convergências e divergências nos padrões de sequências finais.....	100
Quadro 5 - Resultado geral do experimento online: pseudopalavras mais frequentes.....	108
Quadro 6 - Comparação: Dicionário x Corpus Brasileiro x Experimento online	112

SUMÁRIO

1 INTRODUÇÃO	14
1.1 HIPÓTESE E OBJETIVOS.....	16
1.1.2 Hipótese	17
1.1.3 Objetivos	17
2 DELIMITAÇÃO DO TEMA	18
2.1 DERIVAÇÃO REGRESSIVA E IMPREDIZIBILIDADE DA VOGAL FINAL.....	19
2.1.1 Direcionalidade da Derivação Regressiva	21
2.1.2 Generalização da Derivação Regressiva	22
2.1.3 Natureza e Base da Derivação Regressiva	23
2.2 SÍNTESE DOS PROBLEMAS ASSOCIADOS À DISCUSSÃO SOBRE A DERIVAÇÃO REGRESSIVA	26
3 REVISÃO TEÓRICA	29
3.1 PRODUTIVIDADE	29
3.1.1 Aronoff (1976)	30
3.1.2 Basilio (1980)	31
3.1.3 Corbin (1987)	31
3.1.4 Anderson (1992)	32
3.1.5 Baayen (1992)	34
3.1.6 Bauer (2001)	35
3.2 FREQUÊNCIA E MEMÓRIA	37
3.2.1 Fonologia de Uso	37
3.2.2 Teoria dos Exemplos	40
3.3 O STATUS DA VOGAL TEMÁTICA NO PORTUGUÊS BRASILEIRO.....	44
3.3.1 Camara Jr. (1969,1970)	44
3.3.2 Kehdi (1999)	45
3.3.3 Lee (1995)	45
3.3.4 Moreno (1997)	46

3.3.4 Alcântara (2010).....	47
3.3.5 Schwindt (2018).....	50
4 PROCEDIMENTOS METODOLÓGICOS	52
4.1 LÉXICO DICIONARIZADO	52
4.2 LÉXICO INSTITUCIONALIZADO	54
4.3 CODIFICAÇÃO SEGUNDO FATORES LINGUÍSTICOS	56
4.3.1 Vogal temática verbal.....	56
4.3.2 Identidade entre base verbal e nome derivado.....	56
4.3.3 Complexidade morfológica	57
4.3.4 Vogal média tônica.....	57
4.3.5 Gênero da palavra.....	57
4.3.6 Número de sílabas	58
4.3.7 Onset da última sílaba	58
4.3.8 Modo de articulação do contexto fonológico precedente.....	59
4.3.9 Ponto de articulação do contexto fonológico precedente.....	59
4.4 COMPARAÇÃO ENTRE FREQUÊNCIA DE <i>TYPES</i> E <i>TOKENS</i>	60
4.5 PADRÃO DE SEQUÊNCIA SONORA	60
4.6 LÉXICO POTENCIAL: EXPERIMENTO ONLINE COM PSEUDOPALAVRAS	60
4.7 COMPARAÇÃO ENTRE RESULTADOS: LÉXICO DICIONARIZADO, LÉXICO INSTITUCIONALIZADO E LÉXICO POTENCIAL.....	62
5 RESULTADOS	63
5.1 LÉXICO DICIONARIZADO: DICIONÁRIOS ELETRÔNICOS <i>AURÉLIO</i> E <i>HOUAISS</i> ...	63
5.1.1 Datação das palavras	66
5.1.2 Análise de Regionalismos, Formas com Homofonia Acidental, Formas Alternantes e Formas Não Listadas	67
5.1.3 Resultado da codificação segundo os grupos de fatores linguísticos.....	75
5.2 LÉXICO INSTITUCIONALIZADO: <i>CORPUS BRASILEIRO</i>	82
5.2.1 Vogal temática verbal.....	83
5.2.2 Vogal temática	84
5.2.3 Complexidade morfológica	84

5.2.4 Vogal média tônica.....	85
5.2.5 Gênero da palavra.....	85
5.2.6 Número de sílabas.....	86
5.2.7 Onset da última sílaba.....	86
5.2.8 Contexto fonológico precedente (modo de articulação).....	87
5.2.9 Contexto fonológico precedente (ponto de articulação).....	88
5.3 COMPARAÇÃO <i>TYPES</i> E <i>TOKENS</i> : LÉXICO DICIONARIZADO E LÉXICO INSTITUCIONALIZADO.....	88
5.3.1 Vogal temática verbal.....	89
5.3.2 Vogal temática.....	90
5.3.3 Complexidade morfológica.....	90
5.3.4 Vogal média tônica.....	91
5.3.6 Número de sílabas.....	91
5.3.7 Onset da última sílaba.....	92
5.3.8 Contexto fonológico precedente (modo de articulação).....	92
5.3.9 Contexto fonológico precedente (ponto de articulação).....	93
5.3.10 Convergências e divergências.....	93
5.3.11 Padrões de sequências finais.....	94
5.4 EXPERIMENTO ONLINE COM PSEUDOPALAVRAS.....	100
5.4.1 Fatores sociais.....	101
5.4.2 Seis questões de múltipla escolha.....	103
6.5 COMPARAÇÃO ENTRE LÉXICO DICIONARIZADO, LÉXICO INSTITUCIONALIZADO E LÉXICO POTENCIAL.....	111
6 CONSIDERAÇÕES FINAIS.....	114
REFERÊNCIAS.....	122
APÊNDICE A – PADRÕES NAS TERMINAÇÕES DOS SUBSTANTIVOS DEVERBAIS COLETADOS DO DICIONÁRIO ELETRÔNICO AURÉLIO.....	125
APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA O EXPERIMENTO DE PSEUDOPALAVRAS.....	131
APÊNDICE C – EXPERIMENTO COM PSEUDOPALAVRAS.....	132

APÊNDICE D – QUADRO DE CÓDIGOS DOS GRUPOS DE FATORES LINGÜÍSTICOS	134
APÊNDICE E – QUADRO DE RESPOSTAS DO EXPERIMENTO	136
APÊNDICE F – QUADRO DE DISTRIBUIÇÃO DAS RESPOSTAS DOS INFORMANTES AO EXPERIMENTO ONLINE SEGUNDO OS FATORES SOCIAIS	143
APÊNDICE G – LISTA DE FREQUÊNCIA DOS NOMES DEVERBAIS	147
ANEXO A – ÚLTIMO PARECER DO COMITÊ DE ÉTICA EM PESQUISA	163

1 INTRODUÇÃO

A curiosidade pelo tema da produtividade da derivação regressiva nasce de uma pergunta. Num dos meus primeiros anos como professor de Língua Portuguesa, numa turma de nono ano do Ensino Fundamental, ao explicar um conteúdo de morfologia sobre os processos de formação de palavras como sufixação, composição etc., quando apresentava alguns exemplos de derivação regressiva, como *ataque* e *abandono*, ouvi a seguinte pergunta de um aluno: “Por que não é a mesma letra? Por que é *e* em ataque e *o* em abandono?”

Em primeiro lugar, nunca havia pensado sobre aquilo. Em segundo lugar, como estava ainda nos meus primeiros anos como professor, ainda não tinha um “tempo de reação” rápido o suficiente para não deixar um silêncio constrangedor se acumular na sala de aula. Meu raciocínio buscava uma resposta possível, mas realmente não havia meio de explicar de forma satisfatória para aquele aluno o curioso fenômeno que tinha chamado a sua atenção. O momento passou e apenas pude preencher aquele silêncio que ameaçava não se romper com um clichê do tipo “vou estudar sobre isso e te respondo depois”. Muito tempo passou, mas aquela dúvida permaneceu e me impeliu a escolher esse tema para a tese de doutorado. Esta tese é, pois, minha tentativa de explicar para aquele aluno curioso e para mim mesmo por que, a partir dos padrões produtivos da derivação regressiva, às vezes emerge a vogal *o*, como em *abandono*, às vezes a vogal *a*, como em *ajuda*, às vezes a vogal *e*, como em *ataque*.

Nessa perspectiva, neste estudo analisa-se a produtividade do processo morfológico conhecido como derivação regressiva. Este tipo de formação de palavras prevê que, a partir de uma base verbal, forma-se um substantivo (denominado deverbal) que aparenta “reduzir” o material fonológico da forma primitiva (se tomarmos o verbo no infinitivo como a forma primitiva). Assim, para uma forma como *lutar* podemos derivar o deverbal *luta*, para *combater*, *combate*, e para *passar*, *passagem*. O caráter regressivo desse processo parece ser confirmado quando o comparamos a outros processos de nominalização como sufixação com *-mento* e *-ção*, em que, contrastadas base e forma derivada, observa-se um acréscimo de segmentos à forma derivada.

Esta pesquisa tem um objetivo principalmente descritivo e norteia-se pelas noções de **disponibilidade** e **rentabilidade**, e de **armazenamento de exemplares no léxico**. A noção bipartite disponibilidade/rentabilidade, conforme Corbin (1987), assume a existência de duas abordagens possíveis para considerar um fenômeno morfológico como produtivo: a disponibilidade no léxico dicionarizado e a rentabilidade no léxico institucionalizado. Já a

Teoria de Exemplos é uma proposta teórica defendida por Pierrehumbert (2001), que prevê a existência no léxico dos falantes de nuvens de exemplos, agrupamentos de formas semelhantes foneticamente que são determinadas pela frequência de uso de uma forma linguística.

Tomamos como hipótese principal a ideia de que **os padrões produtivos do português brasileiro sinalizam que a derivação regressiva é um processo especial de nominalização, que toma o radical da palavra como base da derivação, acrescentando uma vogal final que é influenciada por fatores linguísticos.**

Em nossa pesquisa analisamos a produtividade da derivação regressiva sob a ótica da vogal final no âmbito da disponibilidade no léxico dicionarizado e, no âmbito da rentabilidade, no que chamamos de léxico institucionalizado e no léxico potencial. Como léxico dicionarizado, entendemos a lista parcial de palavras já fixadas pelo uso no momento da elaboração de um dicionário, sem discriminação da frequência lexical maior ou menor de cada vocábulo listado. Como léxico institucionalizado¹, compreendemos a lista de palavras que são fixadas pelo uso no momento em que a pesquisa acontece, ou seja, o recorte sincrônico dos padrões de uso da língua, com a informação da frequência lexical maior ou menor das palavras listadas a partir de um *corpus* de textos escritos e/ou orais de grande abrangência. Por fim, como léxico potencial entendemos as formas que podem, seguindo os padrões linguísticos da língua, vir a existir, ou que podem já existir apesar de ainda não terem sido registradas nos dicionários.

Assim, a fim de realizarmos uma análise da disponibilidade observada no léxico dicionarizado do português brasileiro (PB) para o fenômeno da derivação regressiva, na primeira etapa de nosso estudo, procedemos à coleta de substantivos deverbiais por meio do *Dicionário Eletrônico Aurélio* versão 5.0.40. Os dados coletados foram codificados a partir de uma série de fatores linguísticos que expressam hipóteses fonológicas e morfológicas que podem estar influenciando a emergência da vogal final, que pode ser *o*, *a* e *e*. Na segunda etapa, no âmbito da rentabilidade, tomamos como representante do léxico institucionalizado o padrão de uso encontrado no *Corpus Brasileiro*, por meio da frequência lexical dos vocábulos já identificados no Dicionário Aurélio na primeira etapa. Na terceira etapa, ainda no âmbito da rentabilidade, investigamos o léxico potencial de falantes nativos do português brasileiro por

¹ Tomamos emprestado o termo “institucionalizado” de Bauer (1988, p. 67), que o utiliza para se referir à “institucionalização” das palavras quando elas passam a ser utilizadas no meio social e, conseqüentemente, listadas nos dicionários. Dessa forma o estágio do léxico institucionalizado pode ser entendido como anterior ao do léxico dicionarizado e posterior ao do léxico potencial, ou seja, um estágio intermediário na formação de novas palavras numa língua natural. Também nos baseamos na perspectiva do lexicólogo Lipka (1992, p. 96), que considera que uma nova forma é institucionalizada quando é reconhecida pelos falantes como um item lexical.

meio de um experimento online, com o uso de pseudopalavras, aplicado a alunos de cursos de graduação de universidades públicas e particulares em território brasileiro, a fim de observar a frequência das três vogais possíveis na derivação regressiva e sua correspondência com as formas infinitivas nas três conjugações. Ao final, comparamos os padrões de produtividade da derivação regressiva no léxico dicionarizado, no léxico institucionalizado e no léxico potencial.

A seguir apresentamos a estrutura da tese.

No Capítulo 2, apresentamos nossa hipótese e nossos objetivos. No Capítulo 3, delimitamos o tema de nossa pesquisa, abordando os principais problemas associados à descrição da produtividade da derivação regressiva. No Capítulo 4, procedemos à revisão teórica referente à Produtividade Morfológica e suas implicações para a Fonologia de Uso, a Teoria dos Exemplos e a descrição da vogal temática no PB. No Capítulo 5, descrevemos nossos procedimentos metodológicos, que incluíram a coleta de nomes derivados via derivação regressiva no léxico dicionarizado, por meio dos dicionários eletrônicos *Aurélio* e *Houaiss*, a codificação desses dados segundo grupos de fatores linguísticos, a análise dos padrões de uso desses nomes derivados no léxico institucionalizado, segundo o *Corpus Brasileiro*², e a verificação da produtividade no léxico potencial por meio da aplicação de um experimento online com pseudopalavras para 100 informantes, alunos de graduação. No Capítulo 6, apresentamos e discutimos os resultados dos padrões produtivos da derivação regressiva sob a ótica da vogal final segundo os grupos linguísticos em análise, comparando os resultados obtidos no léxico dicionarizado, no léxico institucionalizado e no léxico potencial. No Capítulo 7, apresentamos nossas considerações finais, retomando nossa reflexão sobre o tema da produtividade da derivação regressiva sob a ótica da vogal final a partir dos resultados apresentados pela nossa pesquisa; além disso, discutimos alguns problemas residuais que abrem espaço para pesquisas futuras.

A seguir apresentamos nossa hipótese e objetivos.

1.1 HIPÓTESE E OBJETIVOS

Nesta seção apresentamos a hipótese que sustenta nossa análise, seguida dos objetivos geral e específicos que orientaram este trabalho.

² <https://www.linguateca.pt/acesso/corpus.php?corpus=CBRAS>

1.1.2 Hipótese

Este estudo norteia-se pela seguinte hipótese:

Os padrões produtivos sinalizam que a derivação regressiva é um processo especial de nominalização na derivação regressiva que toma o radical³ como base da derivação, acrescentando uma vogal final que é influenciada por fatores linguísticos.

1.1.3 Objetivos

Esta pesquisa tem o seguinte objetivo geral:

Investigar a produtividade da derivação regressiva sob a ótica da vogal final.

Esta pesquisa tem os seguintes objetivos específicos:

1. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico dicionarizado (Dicionários Eletrônicos *Houaiss* e *Aurélio*);
2. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico institucionalizado (*Corpus Brasileiro*);
3. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico potencial (experimento online com pseudopalavras).

³ Em nossa pesquisa entendemos “radical” como “radical primário” ou “raiz”, isto é, numa palavra como *abandono*, nosso radical é constituído da sequência /ABANDON/, a que é acrescida a vogal final *o* durante o processo morfológico da derivação regressiva.

2 DELIMITAÇÃO DO TEMA

Os casos de nominalização que tomam como base formas verbais no português brasileiro abrangem a sufixação, a derivação imprópria e a derivação regressiva.

Para os casos de sufixação, temos exemplos como *destruição* e *condicionamento*; a derivação imprópria refere-se aos casos em que não há alteração da base, mas apenas alteração da categoria lexical, como em “Ele tem um andar muito determinado”, em que do verbo se derivou um substantivo, preservando-se a estrutura morfológica e fonológica da suposta base.

Por outro lado, a derivação regressiva nomeia o fenômeno de aparente redução, morfológica e fonológica, de uma forma verbal derivante quando comparada à sua forma nominal derivada (*lutar/luta*). Para os autores da Gramática Tradicional⁴, há consenso quanto ao fato de haver redução do material fonológico e morfológico da base.

Segundo o que até agora se tem visto, consiste a derivação em se formarem novas palavras acrescentando-se a um radical ora um prefixo (derivação prefixal), ora um sufixo (derivação sufixal), ora um prefixo e um sufixo ao mesmo tempo (derivação parassintética). O termo derivado resulta, pois, da ampliação do termo derivante. Na derivação regressiva ocorre exatamente o oposto: o termo derivado resulta da redução do derivante, por isso que a este se lhe subtrai um segmento terminal. (ROCHA LIMA 2011, p. 266)

Além disso, considera-se a emergência da vogal final como um processo aleatório, já que, aparentemente, não é possível definir quais são contextos linguísticos que determinam a distribuição das vogais finais *o*, como em *abandono*, *a*, como em *fuga*, e *e*, como em *resgate*. Da suposta **impredizibilidade** desse fenômeno morfológico derivam-se outros três problemas que também estão implicados na descrição da produtividade da derivação regressiva:

1. **Direcionalidade:** como podemos ter certeza de que estamos diante de um processo de nominalização que toma como base um verbo e não o contrário?
2. **Generalização:** todos os casos de redução da base são casos de derivação regressiva?
3. **Natureza e base da derivação:** esse tipo de derivação é realmente regressiva, ou seja, há diminuição da palavra? ou estamos diante de um processo de sufixação em que o sufixo é zero? ou, ainda, temos um processo de adição de vogal temática e consequente

⁴ Em nossa pesquisa não nos comprometeremos com a distinção entre Gramática Tradicional, no sentido de estudos gramaticais clássicos com finalidade descritiva, e Gramática Prescritiva ou Normativa, no sentido de prescrição de regras de uso mais adequadas para os falantes de uma língua natural. Assim, neste trabalho, o termo Gramática Tradicional pode assumir tanto o aspecto descritivo quanto o aspecto normativo.

substituição da desinência verbal de infinitivo? A base da derivação regressiva é uma palavra ou um morfema? Temos como base uma forma verbal flexionada ou um radical?

A seguir apresentamos, inicialmente, uma discussão sobre a imprevisibilidade da emergência da vogal final e sua relação com a produtividade da derivação regressiva, e, em seguida, diversas propostas que tentam responder aos questionamentos derivados da questão da imprevisibilidade da derivação regressiva.

2.1 DERIVAÇÃO REGRESSIVA E IMPREVISIBILIDADE DA VOGAL FINAL

Como observamos em formas como *ataque* (atacar) e *beijo* (beijar), não são todos os casos em que há correspondência entre vogal temática do verbo e a do nome derivado. Será que não há como prever em quais condições emergirão as vogais *a*, *e* ou *o*?

Se o processo morfológico da derivação regressiva for um fenômeno produtivo no PB, estamos diante, conseqüentemente, de um processo linguístico previsível, em que a emergência da vogal final é condicionada por fatores linguísticos, ou seja, pode ser definida por uma Regra de Formação de Palavra (que discutiremos mais adiante no capítulo 4). Assim, ao afirmar-se o caráter produtivo desse processo morfológico, rejeita-se sua aparente aleatoriedade.

Said Ali (1931, p. 163) organiza os nomes derivados por derivação regressiva em quatro grupos, segundo o gênero gramatical da palavra: (i) masculino em *o* (*atraso*, *assento*, *emprego*, *voo*, *esforço*, *choro* etc.); masculino em *e* (*embarque*, *desembarque*, *combate*, *corte*, *toque* etc.), feminino em *a* (*amarra*, *pesca*, *sobra*, *suplica*, *leva*, *engorda* etc.) e masculinos e femininos (*pago/paga*, *custo/custa*, *troco/troca*, *grito/grita* etc.). É importante salientar, na organização proposta por Said Ali, o grupo de formas que alternam a vogal final como *custo* e *custa*, que serão abordados nesta tese (ver Capítulo 6).

Segundo Resende (2016, p. 11), não há motivação alguma por trás da seleção da vogal temática em deverbais como *luta*, *ataque* e *encontro*.

[...] uma característica consensualmente reconhecida para a derivação regressiva é a de que a vogal final é idiossincrática, isto é, não há motivações para um nome preferir *a* ao invés de *e* ou *o*, por exemplo; de todo modo, essa é uma propriedade de qualquer nome temático em português. Assim, se a vogal que aparece no final do deverbal regressivo é mesmo uma vogal temática nominal, não haveria razões para se esperar que houvesse coincidência com a vogal temática do verbo. (RESENDE, 2016, p. 11)

Já em Basilio (2011, p. 38-39), podemos encontrar critérios de distinção das ocorrências em *a*, *e* ou *o* segundo a sua frequência e produtividade:

A derivação regressiva com apoio na vogal *a* só ocorre em formações tradicionais, ao contrário da baseada em *o*, utilizada em formações recentes de caráter gíriático[...]. Formações em *e* são também produtivas, sobretudo em verbos prefixados com *re-* e *des-*, nos quais a derivação regressiva é relativamente comum. (BASILIO, 2011, p. 38-39)

Rodrigues (2013b, p. 236) apresenta, para o português europeu, um quadro (Figura 1) que ilustra essa suposta aleatoriedade da emergência da vogal final, em que não se pode basear na vogal temática verbal para prever a vogal final dos nomes deverbais.

Figura 1 - Relação arbitrária entre vogal temática e vogal final

Deverbal	marcador de classe -a	marcador de classe -e	marcador de classe -o
Verbo 1.ª conj.	abraçar > abraço	sacar > saque	recuar > recuo
Verbo 2.ª conj.	recolher > recolha	combater > combate	sorver > sorvo
Verbo 3.ª conj.	zurzir > zurza	curtir > curte	cuspir > cuspo

Fonte: RODRIGUES, 2013b, p. 236

Um modo de testar a suposta imprevisibilidade de emergência da vogal temática é, em primeiro lugar, a consulta a léxicos dicionarizados; em seguida, a consulta a *corpora* de diversos léxicos institucionalizados, de origem oral ou impressa, o que permitiria medir a frequência lexical de palavras com cada uma das três vogais temáticas e cruzá-las com fatores como vogal temática do verbo derivante, número de sílabas do nome deverbal, aspecto lexical do verbo, gênero do nome etc. Essa consulta ao léxico dicionarizado e ao léxico institucionalizado é justificada porque a frequência maior ou menor de determinado padrão disponível na língua e no padrão de uso dos falantes pode ser um forte indício para se deduzir uma regra gramatical da língua. Além disso, a verificação da previsibilidade desse fenômeno pode ser realizada no âmbito do léxico potencial, com experimentos com pseudopalavras, criadas a partir de formas existentes no léxico (seguindo padrões fonológicos, morfológicos, semânticos, sintáticos e pragmáticos), mas dessas diferenciadas o suficiente para não suscitar mera analogia. A partir disso, ao verificarmos se há previsibilidade na derivação regressiva, investigando os fatores que podem ser determinantes para a emergência da vogal final, podemos

descrever a produtividade desse fenômeno no léxico dicionarizado, no léxico institucionalizado e no léxico potencial.

Schwindt (2020, p. 283), em seu estudo sobre a marcação de gênero gramatical no PB, afirma que “a intuição dos falantes que permite predizer em grande medida e de forma acertada o gênero dos nomes se funda em aspectos de produtividade – aqui entendida [...] como a potencialidade de novas palavras serem criadas por um processo morfológico particular”. Assumindo esse ponto de vista, defendemos, neste trabalho, que a predizibilidade (em nosso caso, da vogal final nos derivados regressivos) pode estar estreitamente relacionada à noção de produtividade desse processo.

A seguir revisamos as ideias propostas por alguns autores para lidar com as três questões derivadas da suposta imprevisibilidade da derivação regressiva: a direcionalidade, a generalização e a natureza e base da derivação.

2.1.1 Direcionalidade da Derivação Regressiva

Como saber se um substantivo como *beijo* é derivado do verbo *beijar* ou se é a forma primitiva que serve de base para a formação do verbo? Esse questionamento implica a reflexão sobre os critérios que definem a direcionalidade de um fenômeno morfológico como a derivação regressiva. Para haver derivação regressiva é necessário que a base seja um verbo que forma um nome; para todos os outros casos em que a base é um nome, não estamos diante de um caso de derivação regressiva. Há várias propostas que buscam determinar os casos verdadeiros de derivação regressiva, ou seja, em que há realmente a nominalização de formas verbais.

Do ponto de vista da Gramática Tradicional, há apenas um critério para distinção entre casos falsos e verdadeiros de derivação regressiva: o conteúdo semântico dos nomes.

É preciso não esquecer que os substantivos deverbais denotam “ação”: assim, o *vozeio*, o *embarque*, a *disputa* significam, respectivamente, a ação de *vozear*, a de *embarcar*, a de *disputar*. Já os substantivos que nomeiam “objetos”, ou “substâncias” — estes é que são os primitivos, como *escudo* e *azeite*, que deram origem aos verbos *escudar* e *azeitar*. (ROCHA LIMA 2011, p. 267)

O grande problema desse critério são os casos em que o nome não designa nem uma ação, nem um objeto, como *atraso*, *demora* etc. (LOBATO, 1995).

Basilio (1987, p. 41) propõe que se aplique o critério de se há ou não sentido verbal no uso dos nomes para determinar os autênticos casos de verbos derivando nomes de forma regressiva:

Assim, por exemplo, como *A demora de Maria está aborrecendo Pedro* pode ser interpretada como *Pedro está ficando aborrecido porque Maria está demorando*, *demora* deve ser considerada como uma formação deverbal. Já em *O enfeite de Maria não durou muito* não podemos ter a interpretação *Maria não demorou muito se enfeitando* ou *Maria não demorou muito sendo enfeitada*; conseqüentemente, *enfeite* deve ser considerado como um substantivo básico. (BASILIO, 1987 , p. 41)

Gamarski (1988) estabelece diversos critérios para definição da direção da derivação:

- a) morfológico – há prefixos que só se combinam com bases verbais como *re-* e *des-*; assumindo-se que há nomes com esses prefixos como *reenvio* e *desconto*, podemos considerar esses casos como autênticas derivações regressivas já que não poderiam ter uma base nominal com esses prefixos sem que antes eles tivessem sido derivados pelo verbo prefixado;
- b) semântico – verbos costumam estender seu conteúdo semântico para seus nomes derivados; diante das formas *faro* e *farejo*, percebemos que somente a última pode ter sido derivada do verbo *farejar* já que em *faro* não há reprodução da carga semântica de iteração que vemos *farejo*;
- c) teste sintático – ao construir uma frase com o nome deverbal podemos observar a carga semântica do verbo além de seu possível sentido nominal, se pudermos substituí-lo pelos vocábulos *ato*, *atitude*, *ação* etc.

Desse modo, comparando as propostas acima, podemos perceber que é consenso entre todas elas que o nome deve ser portador do sentido verbal da base da qual deriva. Essa característica aproxima esses casos das nominalizações por sufixos como *-mento* e *-ção*, que podem ser utilizadas tanto com um sentido nominal quanto com um sentido propriamente verbal.

2.1.2 Generalização da Derivação Regressiva

Há uma grande confusão no momento de separar os fenômenos de supressão de elementos de um vocábulo, como *delega* e *chima*, de casos de derivação regressiva. Conforme afirma Basilio (1987, p. 37),

Nos casos de redução ou abreviação, também temos uma palavra formada pela supressão de alguma parte da palavra derivante. Nesses casos, a parte a ser suprimida

é, muitas vezes, imprevisível; e a palavra formada é sinônima da derivante, apenas sendo usada, as mais das vezes, num estilo mais coloquial. Tal é o caso de formas como *boteco*, por *botequim*, *Sampa*, por *São Paulo*, *granfa*, por *grã-fino*, *delega*, por *delegado*. [...] Nos casos de derivação regressiva propriamente dita, a supressão é sempre de uma sequência tônica tomada como afixo: e a palavra resultante não tem o mesmo significado ou uso da palavra derivante. (BASILIO, 1987, p. 37)

Do ponto de vista de estudiosos da Gramática Tradicional, alguns autores confundem a derivação regressiva com casos diacrônicos de truncamento:

[...] consiste em criar palavras por analogia, pela subtração de algum sufixo, dando a falsa impressão de serem vocábulos derivantes: de atrasar tiramos atraso, de embarcar, embarque; de pescar, pesca; de gritar, grito. Assim também os vocábulos *rosmaninho* e *sarampão* foram tomados, respectivamente como diminutivo e aumentativo, marcados, portanto, com sufixos de grau, e daí se tiraram as formas regressivas *rosmano* e *sarampo*, como falsos primitivos. (BECHARA, 2009, p. 430)

[...] a chamada derivação regressiva, que consiste na redução da palavra derivante por uma falsa análise da sua estrutura [...] (CUNHA & CINTRA, 2016, p. 117)

Há critérios que podem ser utilizados para separar fenômenos de redução de casos de derivação regressiva: (i) alteração do significado; (ii) contexto de uso; (iii) processo de reanálise. Em relação ao critério (i), formas como *chimarrão* e *chima* mantêm intacto o conteúdo semântico e só variam no grau de informalidade atribuído à forma reduzida, além de não mudarem a categoria lexical, já que ambas são substantivos. Quanto ao critério (ii), para casos como *fotografia* e *motocicleta*, é muito comum o uso das formas truncadas *foto* e *moto* em contextos mais informais, que não alteram nem a semântica nem a categoria lexical da base. No que concerne ao critério (iii), alguns gramáticos consideram formas como *sarampo* e *rosmano*, como formas que foram originalmente reduzidas de *sarampão* e *rosmaninho*, que foram erroneamente analisadas, por parte dos falantes, como contendo os sufixos *ão* e *inho*, ou seja, como formas derivadas (BECHARA, 2009, p. 400).

Só podemos considerar como casos de derivação regressiva aqueles em que a base verbal sofre um processo de redução e forma um nome terminado em *a*, *o* ou *e*. De um lado temos casos de redução por razões de coloquialidade ou falha na análise morfológica por parte dos falantes, e de outro, temos um processo de nominalização com o mesmo *status* da derivação *sufixal* em *-ção* ou *-mento*, em que se toma uma base verbal para derivar um nome.

2.1.3 Natureza e Base da Derivação Regressiva

Como podemos representar morfológicamente essa regra que deriva nomes de verbos?

Se considerarmos a derivação regressiva um processo de supressão de parte da base verbal, precisaremos definir de que base estamos falando e o que motiva a sua redução. Por outro lado, se procurarmos associar esse processo de nominalização aos casos mais comuns de derivação sufixal como *armamento* e *destruição*, é necessária uma descrição que se oponha ao caráter regressivo desse processo e assumamos que em casos como *luta* ou *ataque* também temos a concatenação de um sufixo.

Além disso, representar morfologicamente o processo acarreta tomar uma importante decisão sobre a que base estamos nos referindo: uma palavra ou um morfema? Um radical a que se adiciona uma vogal temática ou uma forma verbal conjugada a que se acrescenta um sufixo nominal?

Bechara (2009, p. 430), diferentemente de outros autores da Gramática Tradicional, considera como base para os derivados deverbais a primeira ou terceira pessoa do singular do presente do indicativo:

[...] o que explica sua distribuição em substantivos de tema em *o* (se provindos da 1.^a pessoa) ou de tema em *a* ou *e* (se provindos da 3.^a pessoa), sem que se possa prever a opção da norma para a escolha da vogal temática. Os de tema em *o* têm maior vitalidade no português moderno, especialmente na variedade informal: o amasso, o agito, o chego, o sufoco, o apago. (BECHARA, 2009, p. 430)

Para Rocha (2003, p. 187), a derivação regressiva consiste no processo de acréscimo de sufixo, uma “Regra de Formação de Palavra (RFP) de derivação zero”:

$$[\text{patrulhar}]V \rightarrow [[\text{patrulhar}]V \emptyset]N$$

O grande problema dessa análise é não explicar porque não há coincidência entre o tema do verbo e a vogal temática dos nomes em casos como *ataque* (deverbal do verbo *atacar*) e *suborno* (deverbal do verbo *subornar*).

Lobato (1995) propõe uma Regra de Análise Estrutural (RAE) que rejeita a supressão da desinência de infinitivo e toma como base da derivação o radical, ao qual é acrescentada uma vogal temática. Assim, em *luta* temos a seguinte RAE:

$$[[\text{RAD}] + \text{VT}] > [\text{lut-}] + a > [\text{luta}]$$

A vogal temática é afixada ao radical por pressão de uma restrição fonológica que bloqueia que o radical terminado em consoante possa emergir como nome (*lut). Aqui não há propriamente uma derivação, apenas verbo e nome compartilham o radical, o que permite

explicar porque em geral não há coincidência entre a vogal temática do verbo e a do nome: como vemos nas formas *atacar* e *ataque*.

Resende (2016, p. 9) opõe-se à proposta de Lobato, já que, para ele, a vogal temática não emerge apenas por uma restrição fonotática:

[...] a existência de nomes deverbais como *floreio*, *rodeio* etc. constitui um argumento contra a ideia de que a vogal temática só serve para impedir que radicais que violam as restrições fonotáticas do português ocorram como formas livres, já que, ao considerar *rodear*, cujo radical é *rode-*, não haveria motivações (fonológicas ou morfofonológicas) para que a forma derivada (regressiva) fosse *rodeio* e não *rodo*, *rode* ou até mesmo *roda*. Isso porque, sendo *rode-* um radical terminado em vogal, a inserção de uma nova vogal seria desmotivada. O mesmo ocorre com *floreio*, *sorteio*, *passaio*, *bloqueio* etc.

Contudo, o radical em *rodeio* não tem este *e* aludido por Resende (2016, p. 9). Ou a palavra seria atemática, [rodeio-], base para novas formações e também uma palavra pronta, ou seria [rod-], e [-eio] seria a forma resultante de [-ear], terminação verbal que recebe em geral um -i- eufônico nesses casos, entre as vogais *e* e *o*.

O autor prefere descrever esse fenômeno como um processo de sufixação zero que toma como base o tema verbal, ao qual é acrescentado um sufixo zero seguido de uma vogal temática nominal (RESENDE, 2016, p. 10):

- a) [BVfala SDØ VTNa]N → [fala]N
- b) [BVcorta SDØ VTNe]N → [corte]N
- c) [BVbeija SDØ VTNo]N → [beijo]N

Essa proposta tem uma vantagem importante sobre outros modelos ao unificar a representação morfológica dos casos de nominalização no português brasileiro, já que tanto em *construção* como em *corte* temos um processo de sufixação. Por outro lado, há o problema de se lançar mão de construtos teóricos por demais abstratos, que não oferecem materialidade fonológica e que forçam uma uniformidade entre fenômenos morfológicos que não oferecem o mesmo grau de disponibilidade no português brasileiro, já que a derivação por sufixos é muito mais registrada no léxico dicionarizado do que a derivação regressiva (ver Capítulo 6).

Em relação ao português europeu, para Rodrigues (2013), a derivação regressiva é classificada como um processo de conversão (derivação imprópria, nos termos da Gramática Tradicional), por exemplo, quando usamos um verbo como nome ou vice-versa, sem alterar a forma do verbo. A autora propõe que a conversão ocorre no léxico, ou seja, o radical verbal é convertido em radical nominal e depois é acrescentada a vogal temática nominal, que resulta de uma condição fonotática que não permite que formas como “lut” ou “brig” emergjam.

Rodrigues (2013, p. 105) aponta a riqueza semântica, que não é previsível, de formas como *passeio*, que “designa o evento de passear, mas também o local onde se passeia”, e *lixa*, que “designa evento e instrumento”, para justificar a necessidade de situar esse fenômeno no léxico.

Como seria possível formatar na sintaxe todos estes semantismos que muitas vezes variam dentro do mesmo tipo semântico da base? Estas variações e riqueza semânticas só são possíveis no léxico, ou seja, num domínio mental em que se arquitetam esquemas que permitem esta variabilidade.

Para além disso, se esta recategorização ocorresse na sintaxe, seria possível um nome converso de qualquer verbo, pois a sintaxe não está sujeita às exceções a que está a morfologia. [...] este tipo de formação de nomes apresenta muitas restrições de seleção das bases verbais. Significa isto que nem todos os verbos podem dar origem a um nome converso. (RODRIGUES, 2013, p. 105)

A autora também rejeita a hipótese de se tomar a forma infinitiva do verbo como base para a derivação regressiva:

Por que razão, então, tomar o infinitivo como forma de base nos casos em que não ocorre afixação? Se para *moagem*, *arborização*, *envelhecimento* estão disponíveis três formas diferentes da base (radical, tema do presente, tema do particípio, respetivamente), por que não se observa da mesma maneira a formação de *abraço*, *voo*, *desmaio*, etc.? (RODRIGUES, 2013, p. 103)

Assim, para Rodrigues (2013), a base para a derivação dos nomes não é a forma infinitiva do verbo, como é defendido por vários autores, mas o radical do verbo, que já contém todas as idiosincrasias semânticas que condicionarão suas possíveis nominalizações e restrições a novas formações. Uma das grandes desvantagens dessa análise é o seu grau de abstração, por lançar mão de um recurso puramente lexical, ou seja, refere-se a um significado que não tem qualquer expoente fonológico para representá-lo.

2.2 SÍNTESE DOS PROBLEMAS ASSOCIADOS À DISCUSSÃO SOBRE A DERIVAÇÃO REGRESSIVA

A partir das diversas propostas expostas acima que abordam o fenômeno morfológico da derivação regressiva, apresentamos a seguir as noções que orientaram nosso estudo.

O que procuramos propor em nossa análise desse processo morfológico é uma alternativa que leve em consideração as particularidades da derivação regressiva, separando-a de processos de formação sufixal e, ao mesmo tempo, contemplando o caráter duplo de sua significação, tanto nominal quanto verbal.

A derivação regressiva apresenta um grande nó complicador em suas possíveis descrições. Por um lado, pode-se considerar, como assumido pela Gramática Tradicional, que esse é um processo de subtração de desinência verbal, em que se toma o infinitivo como base para a derivação, e em que sua forma derivada apresenta o acréscimo de uma vogal temática nominal, que pode ser *o*, *a* ou *e*.

Um dos grandes problemas que os críticos apontam sobre essa descrição do fenômeno é o fato de tomarmos a forma infinitiva como base da derivação (RODRIGUES, 2013). Nesse caso, não estaríamos confundindo a forma listada no dicionário com a forma primitiva? Isso não conduz a um nó explicativo nos casos em que não há coincidência entre vogal temática do verbo e vogal final da forma derivada? Por exemplo, numa forma como *abandonar*, temos a forma derivada *abandono*, em que não se pode recuperar a vogal temática verbal -a-, já que a vogal final da forma derivada é *o*. O fato de se tomar um verbo como base não significa que seja necessário escolher a forma infinitiva, pois há vários fenômenos que podem tomar a forma flexionada como base. Em outras palavras, não seria mais fácil considerar, como propõe o gramático Bechara (2009, p. 430; ver seção 3.3), o substantivo *abandono* como derivado da primeira pessoa do singular do presente do indicativo do verbo *abandonar* (*abandono*)? Se esse fosse caso, substituiríamos o “problema da forma infinitiva” pelo problema da forma flexionada, já que como poderíamos justificar o fato de *abandono* selecionar como base a primeira pessoa do presente do indicativo do verbo *abandonar* e *ataque* ser derivado da primeira pessoa do presente do subjuntivo do verbo *atacar*, e um deverbal como *perda* não se referir a nenhuma forma verbal flexionada conhecida?

Por outro lado, se quisermos incluir esse fenômeno nos casos de nominalização em que há acréscimo de sufixo, como *-mento* e *-ção*, precisamos postular a existência de um “sufixo zero”, que não se manifesta na forma derivada. E ainda podemos incluir esse fenômeno nos casos de derivação por conversão, como por exemplo na frase “O cantar é um ato terapêutico”, em que o verbo *cantar* foi convertido em substantivo sem que houvesse alteração de seu material morfofonológico, ou seja, num caso que constitui uma operação puramente sintática.

Desse modo, percebe-se que se pode dividir esse fenômeno em dois tipos de análise: há os que entendem a derivação regressiva como um **processo especial de nominalização em que não há acréscimo de sufixo** e há os que defendem a ideia de que o processo está incluído nos **casos de nominalização como -mento e -ção**, o que impõe a postulação de um sufixo zero à base, ou **nos casos de conversão**, em que não há alteração do material morfofonológico, apenas alteração de classe no nível sintático. Em suma, ou estamos diante de um fenômeno restrito do português brasileiro, que o aproxima de outros casos de truncamento, mas que em geral não

mudam a classe da palavra reduzida; ou nos vemos diante de um subgrupo de um fenômeno mais geral de nominalização por acréscimo de sufixo ou conversão.

Porque acreditamos que a derivação regressiva é determinada por fatores como disponibilidade no léxico e rentabilidade no uso (CORBIN, 1987) e nuvens de exemplares (PIERREHUMBERT, 2001) no que tange à emergência da vogal final, neste estudo defendemos que **os padrões produtivos do português brasileiro sinalizam que a derivação regressiva é um processo especial de nominalização, que toma o radical da palavra como base da derivação, acrescentando uma vogal final que é condicionada por fatores linguísticos.**

Sob a perspectiva da Teoria de Exemplos, a frequência de uso determina o quanto uma forma é imediatamente acessível à memória do falante. Já do ponto de vista das restrições gramaticais no âmbito da semântica, da morfologia e da fonologia, e até mesmo da pragmática, a base selecionada para essa RFP (Regra de Formação de Palavra) depende do quanto essas restrições podem ser limitadoras para garantir a produtividade ou não dessa regra morfológica.

Isso nos coloca diante de um binômio que orienta o desenho de nossa pesquisa: léxico e uso, ou conforme Corbin (1987), disponibilidade e rentabilidade. Como disponibilidade entende-se a propriedade de uma regra estar contemplada no léxico do falante em maior ou menor grau. Já rentabilidade refere-se à característica de maior ou menor uso de uma regra morfológica. Assim uma regra pode ter um alto grau de disponibilidade no léxico mas, devido a fatores linguísticos e extralinguísticos não ter muita rentabilidade no uso (é quase como o caso do binômio RFP/RAE (Regra de Análise Estrutural), em que mesmo transparente para os falantes, algumas regras de formação de palavras não são mais produtivas, ou seja, não são mais usadas pelos falantes para formar novas palavras). Esses dois conceitos alinham-se às categorias da fonologia de uso de *types* e *tokens*. Dessa forma, disponibilidade pode ser associada ao léxico e a *types*, enquanto rentabilidade, à frequência e a *tokens*.

3 REVISÃO TEÓRICA

Neste capítulo, apresentamos os pressupostos teóricos que nortearam nossa pesquisa. Inicialmente são apresentadas diversas perspectivas que compõem os estudos sobre o conceito de produtividade na Morfologia. Em seguida, debruçamo-nos sobre o aspecto da produtividade que abordamos em nossa análise da derivação regressiva: a frequência de uso dos padrões morfológicos. Para isso, sintetizamos as principais noções que sustentam a Fonologia de Uso e a Teoria dos Exemplares. Ao final do capítulo, discutimos o eixo que norteou nossa análise da produtividade da derivação regressiva: a vogal temática. Com esse propósito, descrevemos algumas propostas teóricas que abordam o *status* da vogal temática no PB.

3.1 PRODUTIVIDADE

Antes de discutirmos a representação lexical dos nomes derivados por derivação regressiva, precisamos refletir sobre a noção de produtividade morfológica. Conforme Spencer (1991, p. 48-49), a discussão sobre a natureza do léxico relaciona-se ao conceito de produtividade:

Se observamos os recursos de formação de palavras da maioria das línguas, percebemos que alguns destes são regular e ativamente usados na criação de palavras totalmente novas, enquanto outros caíram em desuso com a passagem do tempo ou são empréstimos apenas usados em circunstâncias restritas⁵.

Para nos debruçarmos sobre esse tema tão complexo, apresentaremos um breve panorama de diferentes concepções do conceito de produtividade. Assim, neste capítulo nos detemos em várias abordagens do tema produtividade desde Aronoff (1976) e sua proposta de Regras de Formação de Palavra (RFPs), passando por Basilio (1980), que acrescenta à RFP de Aronoff a Regra de Análise Estrutural (RAE), também nos referimos a Corbin (1987), que sugere dois aspectos essenciais para a distinção dos processos produtivos: rentabilidade e disponibilidade. Além disso, apresentamos algumas ideias de Anderson (1992) que questionam os princípios que subjazem ao modelo gerativo seguido pela teoria morfológica. Ao fim, temos a reflexão de Bauer (2001) sobre as várias possibilidades de abordagem do conceito de produtividade e sua proposta particular de definição.

⁵ If we look at the word formation resources of most languages we find that some of these are regularly and actively used in the creation of totally new words, while others have fallen into desuetude with the passage of time, or have been borrowed from elsewhere and are only used in restricted circumstances.

3.1.1 Aronoff (1976)

Aronoff (1976) propõe um modelo gerativo para a derivação morfológica⁶ que toma como base a palavra e não o morfema para a descrição de regras de formação de palavras (WFR, no original em inglês; RFP, em português). A principal implicação dessa noção é a de que no léxico não são listados morfemas, apenas palavras.

Segundo Aronoff (1976, p. 36), uma forma bastante simples de medir a produtividade de determinada regra morfológica seria por meio de uma abordagem quantitativa: “se quisermos comparar a produtividade de duas RFPs, podemos simplesmente fazer listas de palavras formadas pelos respectivos processos e adicioná-los. Quanto mais longa a lista, mais produtiva a RFP”⁷.

Esse método, contudo, conforme apontado pelo autor, não leva em conta restrições morfológicas em relação à base e à classe de palavra envolvidas em determinada regra de formação de palavras (RFP). Essas restrições poderiam ser levadas em consideração se

[...] contamos o número de palavras que sentimos que poderiam ocorrer como output de uma dada RFP (o que podemos fazer contando o número de bases possíveis para a regra), contamos o número de palavras que realmente são formadas por essa regra, tomamos a razão entre os dois, e comparamos esta com a mesma proporção para a razão de outra RFP. De fato, por este método, podemos chegar a um índice simples de produtividade para cada RFP: da razão de palavras possíveis àquelas realmente listadas. (ARONOFF, 1976, p. 36)⁸

Aronoff identifica dois problemas derivados desse método: (i) não é possível falar de produtividade de uma RFP em termos absolutos, somente em relação a uma classe morfológica particular; e (ii) a ideia de quantificar a produtividade implica a noção de listar no léxico cada nova palavra formada, o que parece ser contraintuitivo para o caso de algumas RFPs bastante produtivas.

Aronoff propõe que há uma relação causal entre a noção de léxico e a de produtividade: “a entrada do output de uma RFP no léxico leva a uma perda de produtividade. Quase por acaso,

⁶ Para Aronoff (1976), a flexão e a composição devem ser consideradas no âmbito sintático.

⁷ [...] if we want to compare the productivity of two WFRs, we may simply make lists of the words formed by the respective processes and add them up. The longer the list, the more productive the WFR. (ARONOFF, 1976, p. 36).

⁸ we count up the number of words which we feel could occur as the output of a given WFR (which we can do by counting the number of possible bases for the rule), count up the number of actually occurring words formed by that rule, take a ratio of the two, and compare this with the same ratio for another WFR. In fact, by this method we could arrive at a simple index of productivity for every WFR: the ratio of possible to actually listed words. (ARONOFF, 1976, p. 36).

este segundo ponto responde a uma pergunta colocada no início [...]: todas as palavras novas entram no léxico? A resposta é não”⁹ (ARONOFF, 1976, p. 45).

Assim palavras formadas por RFP só são lexicalizadas quando a composicionalidade das palavras derivadas perde sua transparência semântica.

3.1.2 Basilio (1980)

Basilio (1980) destaca os princípios da gramática gerativa chomskiana envolvidos nas operações morfológicas realizadas pelo falante de uma língua nativa: o conhecimento lexical do falante no armazenamento de uma lista de entradas lexicais, a formação de palavras a partir de regras morfológicas (RFP) e a análise da estrutura de entradas lexicais morfológicamente complexas, resultantes de RFPs que são mais usadas pelos falantes de uma comunidade.

A autora formaliza a RFP da seguinte forma: “Uma RFP (Regra de Formação de Palavras) sufixal é descrita basicamente a partir do esquema abreviado abaixo: [X]A -> [[X]A Y]R, onde X representa a base e A e B são categorias lexicais ou classes de palavras” (BASILIO, 1987, p. 55).

Segundo essa proposta, espera-se que uma RFP no âmbito da produção do falante tenha como contraparte uma Regra de Análise Estrutural (RAE) no domínio da percepção; já uma RAE não implica a existência de uma RFP correspondente. Essa regra distingue percepção de produção, o que significa dizer que regras de formação de palavra que não são mais produtivas ainda podem ser transparentes para o falante.

3.1.3 Corbin (1987)

Para a autora (CORBIN, 1987, p. 177), o conceito de produtividade não pode ser considerado unidimensionalmente, apenas se considerando o uso (rentabilidade) em maior ou menor grau de determinada regra morfológica pelos falantes:

A produtividade de uma formação morfológica, no sentido de que esse termo designa a possibilidade de os falantes terem uma regra de derivação sem outras restrições além daquelas impostas pela própria regra, não pode ser medida pela rentabilidade

⁹ [...] The listing of the output of a WFR in the lexicon leads to a loss in productivity. Almost incidentally, this second point answers a question posed at the very beginning [...] : Are all new words entered in the lexicon? The answer is no” (ARONOFF, 1976, p. 45).

discursiva das palavras construídas de acordo com essa formação, apesar do alcance que essa crença amplamente difundida possa ter. (CORBIN, 1987, p. 177)¹⁰

Corbin (1987, p. 177) apresenta uma distinção importante no que se refere aos processos morfológicos associados à produtividade: há casos em que estamos diante do aspecto da **rentabilidade**, ou seja, uma regra ou processo morfológico resulta em um grande número de novas palavras; e casos em que observamos a **disponibilidade**, em outras palavras, quando um processo ou regra morfológica é oferecido pelo léxico da língua para ser utilizado na formação de novas palavras.

Conforme a autora, “A lógica da rentabilidade é discursiva e depende de variações nos inventários, lexicográficos ou de outra natureza, sujeitos a flutuações que demonstraram ter pouca relação com uma pseudo-evolução do léxico”¹¹ (CORBIN, 1987, p. 177). Por outro lado, para Corbin (1987, p. 177), “A lógica da disponibilidade é linguística e não tem suporte empírico, exceto as manifestações de competência lexical”¹², ou seja, só temos indícios da disponibilidade de determinada regra gramatical por meio de inferências sobre o léxico dos falantes.

Portanto, o conceito de produtividade pode se referir tanto ao caráter quantitativo quanto ao seu potencial para formação de novas palavras. Geralmente tendemos a assumir apenas uma ou outra perspectiva, porém é necessário considerar a produtividade de determinado processo ou regra morfológica a partir dessas duas perspectivas, já que processos que envolvem a diacronia, ou seja as mudanças que ocorrem na língua no decorrer do tempo, podem não só considerar a frequência de um fenômeno mas também a sua potencialidade para implementar transformações na estrutura da língua.

3.1.4 Anderson (1992)

Ao modelo gerativo subjaz o princípio de não redundância, que recomenda que as regras e representações que compõe a gramática de uma língua natural não repitam informações, ou

¹⁰ La productivité d'une formation morphologique, au sens où ce terme désigne la possibilité pour des locuteurs de disposer d'une règle dérivationnelle sans autres contraintes que celles imposées par la règle elle-même, ne peut pas se mesurer à la rentabilité discursive des mots construits dépendant de cette formation, en dépit de l'écho que peut avoir cette croyance, largement répandue.

¹¹ La logique de la rentabilité est d'ordre discursif, et s'appuie sur les variations des inventaires, lexicographiques ou autres, eux-mêmes soumis à des fluctuations dont il a été démontré qu'elles n'ont que peu de rapports avec une pseudo-évolution du lexique

¹² La logique de la disponibilité est d'ordre linguistique, et n'a pour point d'appui empirique que les manifestations de la compétence lexicale

seja, que as informações contidas no léxico (nível morfológico) não sejam repetidas no componente gramatical (nível sintático). Em outras palavras, se existe uma operação sintática que dá conta de determinado output da língua, não é necessário que esse output seja um item lexical; temos de um lado um componente linguístico cujo propósito é estocar tudo que não pode ser previsto por regras gramaticais e de outro, um nível que opera com essas regras gramaticais. Anderson (1992) opõe-se a esse princípio, que está implicado na forma como é concebida a produtividade morfológica. A relação entre regras produtivas e formas não lexicalizadas que se depreende do princípio de não redundância na descrição do léxico pode não ter uma base empírica que a sustente conforme sugere Anderson (1992, p. 195). Conforme aludido na seção 3.1.2, a Teoria dos Exemplares vai além desse posicionamento de Anderson e assume uma redundância ilimitada, já que a informação gramatical é repetida nas nuvens de exemplares, não havendo uma fronteira tão distinta entre regra (gramática) e memória (léxico).

A questão que deriva dessa discussão sobre representação da redundância no léxico ou não diz respeito ao fato de que se deve ou não incluir no léxico palavras formadas por RFPs que são produtivas na língua junto com vocábulos que não podem ser preditos por essas mesmas RFPs produtivas.

Conforme Anderson (1992, p. 196),

A noção de produtividade frequentemente é atraente no sentido do puro número de formas relacionadas a uma dada regra, mas esse certamente não é um parâmetro significativo da estrutura linguística. O que nos interessa não é como muitas formas existem, mas em vez disso quantas formas que uma regra pode prever que existam são possibilidades reais.¹³

Assim, para o autor o que importa é discutir porque algumas formas possíveis, segundo dada RFP, sofrem bloqueio em determinada língua, cuja natureza das restrições pode ser de ordem fonológica, semântica etc., “mas permanecem palavras preditas, cuja inexistência parece arbitrária. Como devemos bloquear, por exemplo, a *scrupulosity* de *scrupulous* sem apelar para um recurso arbitrário”¹⁴ (ARONOFF, 1992, p. 197). Daí surge a noção de “princípio de bloqueio”, que prevê que, caso haja uma RFP produtiva na gramática da língua, será bloqueada qualquer outra RFP que resulte numa forma com conteúdo semântico análogo à primeira.

¹³ The notion of productivity is often appealed to in the sense of the sheer number of forms accounted for by a given rule, but this is surely not a significant parameter of linguistic structure. What interests us is not how many forms exist, but rather how many of the ones a rule might predict to exist are real possibilities.

¹⁴ but there remain words which are predicted but whose nonexistence seems arbitrary. How are we to block, for example, *scrupulosity* from *scrupulous* without an appeal to an arbitrary feature

3.1.5 Baayen (1992)

Contrastando seu estudo com abordagens qualitativas da produtividade morfológica (ver ARONOFF, 1976; BOOIJ, 1977, VAN MARLE, 1985; COSERIU, 1970, 1975; BURGSCHMIDT, 1977), Baayen (1992) tem como objetivo analisar quantitativamente os processos morfológicos produtivos. Esse índice da produtividade (BAAYEN, 1992, p. 115) é obtido por meio da razão entre a frequência de tipo de um processo morfológico e o total de *types* que ocorrem apenas uma vez naquele processo morfológico (hapax legomena):

$$P=n_1/N$$

Esse índice reflete a noção de que processos produtivos geram mais palavras novas, ainda não cristalizadas no léxico do que é observado com processos não produtivos.

Conforme Quadros (2009, p. 24), o índice de produtividade proposto por Baayen demonstra como a transparência semântica atua em processos altamente produtivos, nos quais o significado do vocábulo derivado pode ser predito pela combinação de sentidos da base e do afixo (por exemplo, *re+escrever* – ‘escrever de novo’). Por outro lado, formas morfológicamente complexas, ou seja, semanticamente opacas, têm um sentido não predizível composicionalmente.

Para Baayen (1992, p. 110-111) a análise da relação entre a frequência e a produtividade não pode se limitar a frequência de *types*, em que processos morfológicos com mais *types* são considerados os mais produtivos, já que qualquer medida de produtividade morfológica deve satisfazer as seguintes exigências:

- correspondência entre o ranqueamento de processos de formação de palavras apresentado por uma análise quantitativa e a intuição dos falantes;
- descrição estatística da “prontidão com que um elemento é introduzido em novas combinações” (BOLINGER, 1948, p. 18);
- efeito de diminuição da medida de produtividade em itens com conteúdos semanticamente idiossincráticos;
- complexidade da relação entre produtividade e frequência, que deve ir além do simples levantamento da frequência de *types*.

Ademais, Baayen (1992, p. 125) também aponta para a necessidade de considerarmos o papel da frequência no léxico mental dos falantes no que se refere à atribuição de um índice

alto ou baixo de produtividade para determinado processo morfológico: “a frequência de palavras representa uma instância particular da informação de frequência que é de forma não intencional acumulada na memória, e seu efeito em vários experimentos é bastante conhecido”.¹⁵

3.1.6 Bauer (2001)

Conforme Bauer (2001, p. 2), é consenso entre linguistas que novas combinações de sentenças são criadas a partir de um número finito de elementos estocados aplicando-se regras às quais os falantes não necessariamente precisam ter acesso para as utilizar. A criação de novas combinações é um fenômeno trivial das línguas do mundo que não se limita ao domínio sintático mas também se relaciona à noção de produtividade morfológica.¹⁶

Sobre a discussão sobre o lugar da produtividade no desenho da gramática, Bauer (2001, p. 5) critica a tendência descritiva generalizante de associar marcas linguísticas produtivas a regras e listar marcas não produtivas no léxico; casos como a flexão de plural em holandês demonstram que essa visão é muito simplista e não reflete a realidade de muitas línguas naturais.¹⁷ Por exemplo, em holandês, a marca default de plural é -en, mas há outra marca produtiva de plural, -s (que é determinado por algumas restrições fonológicas, como o tipo de vogal final da raiz, e morfológicas, como o tipo de sufixos adjungidos à base).

O autor (BAUER, 2001, p. 34) comenta que, se nos basearmos na noção de um conjunto de palavras potenciais/possíveis/novas como medida da produtividade de um processo morfológico, devemos ter bem clara a noção do que é, ou os critérios para definir, uma palavra real ou existente na língua.¹⁸

¹⁵ Word frequency represents a particular instance of frequency information that is unintentionally accumulated in memory, and its effect in various experimental tasks is well known. (BAAYEN, 1992, p. 125).

¹⁶ The general assumption among linguists is that people can say things which they have neither said nor heard before because they know (albeit not explicitly) the rules for the combination of appropriate elements into sentences and because they are able to assign meanings to those elements. In other words, starting from a store of elements each of which has a predetermined semantic and/or pragmatic value, people are able to create new combinations which allow the transfer of more complex meanings to their interlocutors. It is the creation of the new combinations from a set of stored elements which is crucial here. (BAUER, 2001, p. 2)

¹⁷ If we accept, for the sake of the argument, the simplistic view of English plural formation outlined above, we might wish to conclude that plural markers are divided into two classes in English: the productive and the non-productive. We might consider that the productive marker was assigned by rule, while the unproductive markers were assigned by lexical list. This general picture is widely accepted in the literature, for example in much work on Lexical Phonology and Morphology. If this simplified picture were true, it might illustrate an ideal case. However, there are other patterns which depart considerably from this ideal. Consider next the case of plural in Dutch. (BAUER, 2001, p. 5)

¹⁸ Van Santen (1992: 63–74) stresses the importance of the notion of ‘actual’ (Aronoff 1976; 1983), ‘occurring’ (Allen 1978: 25) or ‘existing’ (Bauer 1988: 62–64) words for the concept of morphological productivity. If productivity is concerned with the potentiality of new formations, then it must be possible to discover whether or not something is new, and this implies that it can be compared with a list of formations which are not new but

Há uma distinção muito importante proposta por Bauer (2001, p. 38) para os casos de palavras “cunhadas”, que são associadas à produtividade de determinada regra de formação de palavras e os casos de neologismo: palavras cunhadas não se tornam parte da norma de uma comunidade de falantes de uma língua natural, já neologismos se integram ao léxico dessas comunidades. Portanto, conforme Bauer (2001, p. 41):

Produtividade se refere a potencial. Um processo é produtivo se tem potencial para levar a novas palavras sendo cunhadas. Nós temos consciência da produtividade apenas por meio de novas palavras cunhadas e dos padrões de palavras familiares e não familiares cunhadas pelo processo em análise.¹⁹

Também analisa a relação entre frequência e produtividade, que para o autor é uma relação indireta: frequência de *types*, por exemplo, marca uma produtividade passada mais do que uma indicação de produtividade atual (BAUER, 2001, p. 48-49). Assim para definir um processo ou regra como produtivo temos que ir além do aspecto quantitativo; precisamos entender os processos diacrônicos que provocaram mudanças significativas na estrutura da língua.

Sintetizando todo um percurso analítico sobre as implicações de associações de sentido entre produtividade e outros termos como frequência, transparência semântica e fonológica, não marcação, etc., Bauer (2001, p. 98) propõe a seguinte definição: “[...] a produtividade de um processo morfológico é o seu potencial para ser cunhado de forma repetitiva não criativa (ou seja, sem se desviar das regras disponíveis em uma dada língua natural)”²⁰.

Aqui a ideia de repetição parece ser uma noção que precisa ser destacada, já que está intimamente relacionada ao conceito de rentabilidade e à ideia de frequência, conforme defendida pela Fonologia de Uso e pela Teoria dos Exemplares. Produtividade, repetição, rentabilidade, frequência são conceitos que se relacionam essencialmente a um elemento que será muito caro em nosso estudo: os padrões morfológicos de uso da língua.

‘established’ (Bauer 1983: 50). Van Santen’s point here seems to me to be unassailable. If, as she so nicely phrases it, ‘Productivity is manifested in the space between the existing and the impossible’ (Van Santen 1992: 72; My translation, LB), then we must be able to recognize the existing. In brief, if we do not accept a category of existing words, we can have no notion of productivity. (BAUER, 2001, p. 34)

¹⁹ Productivity is all about potential. A process is productive if it has the potential to lead to new coinages, or to the extent to which it does lead to new coinages. We are aware of productivity only through the new coinages and the patterns of familiar and unfamiliar words coined by the relevant process. (BAUER, 2001, p. 41).

²⁰ In sum, the productivity of a morphological process is its potential for repetitive non-creative morphological coining. (BAUER, 2001, p. 98).

3.2 FREQUÊNCIA E MEMÓRIA

Nesta seção discutimos a relação entre a estrutura linguística e o uso, mais especificamente, o papel que a repetição tem na emergência de categorias e regras segundo os modelos da Fonologia de Uso e da Teoria dos Exemplos.

3.2.1 Fonologia de Uso

A Fonologia de Uso (BYBEE, 2001, 2002, 2010) é um modelo teórico que toma o uso da língua como base para a descrição e explicação dos fenômenos linguísticos. Opondo-se a teorias linguísticas formais que adotam como princípio a dicotomia saussureana língua/fala ou a dicotomia chomskiana competência/performance, é reconhecido o caráter essencialmente emergente da estrutura linguística, ou seja, não há *a priori* categorias e regras que orientem os processos linguísticos; essas regras e categorias derivam das relações que vão sendo estabelecidas entre as formas da língua.

Esse tipo de sistema, geralmente chamado de sistema emergente, tem as propriedades que achamos descritas como sistemas complexos na natureza. Em sistemas complexos, um pequeno número de mecanismos opera em tempo real e com a repetição leva ao surgimento do que parece ser uma estrutura organizada, como uma duna de areia. No entanto, sabemos que uma duna de areia não é fixa no tempo e no espaço, mas está sempre mudando e tomando uma nova forma. Portanto, vemos que a linguagem também está sempre em um processo de tornar-se outra coisa - criando, perdendo e recriando estruturas que nunca são absolutamente fixas, permitindo variação e mudança contínuas [...].²¹ (BYBEE, 2007, p. 8)

Além disso, pressupõe-se que não há um conhecimento linguístico específico na mente dos falantes, ou seja, os fatos da língua são resultado de processos cognitivos gerais; assim, segundo essa proposta teórica, os mesmos processos envolvidos no aprendizado de um instrumento musical são utilizados no momento em que a criança está dando seus primeiros passos como usuária de sua língua materna.

Conforme a Fonologia de Uso, o principal fator para entender como se dá a construção do conhecimento linguístico é a **repetição**. Por meio desse processo, dá-se um fortalecimento das representações lexicais.

²¹ This *type* of system, often referred to as an emergent system, has the properties we find described as complex systems in nature. In complex systems, a small number of mechanisms operate in real time and with repetition lead to the emergence of what appears to be an organized structure, such as a sand dune. However, we know that a sand dune is not fixed in time and space but is ever altering and becoming. So we see that language is also always in a process of becoming—creating, losing, and re-creating structures that are never absolutely fixed, allowing for continued variation and change [...].(BYBEE, 2007, p. 8)

Por meio da repetição, obtemos fortalecimento lexical – representações fortes e de fácil acesso, como cumprimentos quando vemos alguém que conhecemos ou respostas como “muito obrigado” e “de nada”; ou seja, qualquer tipo de resposta automática que foi aprendida. É a repetição que ritualiza essas respostas e as torna prontamente acessíveis. Esses são apenas exemplos extremos de um fenômeno geral que permeia a representação linguística –, repetição leva ao fortalecimento da representação. (BYBEE, 2001, p. 8)²²

As outras consequências linguísticas da repetição são a **redução da forma e do significado**, e a **emancipação da palavra**.

Bybee (2001, p. 9) descreve a redução da forma como resultado de um mecanismo de **automação**, em que, quanto mais frequente for a ocorrência de uma palavra, mais propensa ela estará a sofrer redução de seu material fonético. O português brasileiro apresenta vários exemplos dessa redução em sua história, como o caso da palavra *você*, que é resultado de um longo processo de redução da expressão *Vossa Mercê* (NASCENTES, 1956).

Vossa Mercê > Vossemecê > Vosmecê > Você

A redução do significado está relacionada ao processo de **habituação**, que se refere a “um declínio na tendência de responder a estímulos que se tornaram familiares devido à exposição repetida ou persistente” (HAIMAN, 1994, p. 7 apud BYBEE, 2001, p. 9). A autora cita o caso do francês, cuja construção negativa *nes...pas* tinha o *pas* como um elemento enfático, que não carregava o conteúdo negativo, mas agora é parte gramatical obrigatória das frases de negação em francês.

Non, je ne travaille pas. (‘Não, eu não trabalho.’)

*Non, je ne travaille. (forma agramatical)

Como emancipação de uma palavra, Bybee se refere aos casos de **gramaticalização**, em que a palavra é descolada de seu significado instrumental e adquire uma função puramente comunicativa.

(...) a repetição leva à emancipação. Na emancipação, ações instrumentais são dissociadas de sua motivação original e estão livres para assumir uma função

²² Through repetition we get lexical strength – strong, easily accessible representations, such as a greeting when you see someone you know or responses such as ‘thank you’ and ‘you’re welcome’; that is, any kind of learned automatic response. It is repetition that ritualizes these responses and makes them readily available. These are just extreme examples of a general phenomenon that pervades linguistic representation – repetition leads to strength of representation.

comunicativa. A saudação militar deriva de um gesto mais instrumental utilizado na Idade Média quando os cavaleiros em armadura se cumprimentavam. Eles erguiam a viseira do elmo para mostrar suas faces em uma indicação de saudação amistosa. A armadura e a viseira já não existem, mas a forma reduzida do gesto permanece, mesmo que sem sua função instrumental. O gesto não serve mais para erguer a viseira, mas, em vez disso, foi imbuído com a função de comunicar respeito à hierarquia militar. (BYBEE, 2001, p. 8)²³

Por exemplo, no português brasileiro, a palavra *tipo* passou a apresentar outra função e significação em relação a seu uso inicial (BITTENCOURT, 1999), quando usada na expressão coloquial “tipo (assim)”, assumindo a função de um operador discursivo como o advérbio de modo *assim*.

Em outras palavras, a frequência – alta ou baixa – de uma palavra determina sua representação linguística e os processos a que pode ser submetida, como redução de forma e conteúdo para palavras mais frequentes e analogia para palavras menos frequentes. Para palavras mais frequentes, o processo de redução é “resultado da automação da produção linguística” (BYBEE, 2002, p. 267). Já o processo de analogia diz respeito aos casos de palavras menos frequentes que exibem inicialmente um padrão irregular mas que, como não estão prontamente acessíveis à memória, sofrem regularização conforme o padrão das palavras mais frequentes; essas palavras só manteriam seu padrão irregular se apresentassem uma alta frequência de ocorrência (BYBEE, 2002, p. 269-270).

A frequência pode ser observada de duas maneiras: a **frequência de ocorrência** (*token*) e **frequência de tipo** (*type*). Bybee (2001, p. 10) explica que frequência de ocorrência é o número de ocorrências de uma palavra em particular e a frequência de tipo se refere ao número de ocorrências de um determinado padrão da língua (padrão acentual, afixo ou encontro consonantal). Bybee (2001, p. 12-13) ainda argumenta que a frequência de tipo revela a produtividade de determinado padrão da língua.

Parece que a produtividade de um padrão, expressada em um esquema, é enormemente, porém não inteiramente, determinada por sua frequência de tipo: quanto mais itens são abarcados por um esquema, mais forte ele é e mais disponível ele está para ser aplicado a novos itens. Logo, o sufixo de passado do inglês -ed aplica-se a milhares de verbos e é mais produtivo que qualquer dos padrões irregulares, os

²³ (...) repetition leads to emancipation. In emancipation, instrumental actions are disassociated from their original motivation and are free to take on a communicative function instead. The military salute derives from the more instrumental gesture used in the Middle Ages when knights in armor greeted one another. They raised the visor of their helmet to show their faces as an indication of a peaceful greeting. The armor is gone, the visor is gone, but a reduced form of the gesture remains, though without its instrumental function. It no longer raises the visor, but it has been imbued instead with the function of communicating respect for the military hierarchy.

quais são mais restritos em relação ao número de verbos a que se aplicam. (BYBEE, 2001, p. 12-13)²⁴

Desse modo, a Fonologia de Uso analisa os fenômenos linguísticos decorrentes da frequência de uso, ou seja, da repetição de palavras e padrões linguísticos numa comunidade de falantes. O Quadro 1 abaixo resume os tipos de padrões analisados por essa proposta teórica.

Quadro 1 - Fenômenos linguísticos associados à frequência das palavras

Alta frequência	Baixa frequência
Redução da forma - Automação	Analogia
Redução do significado - Habituação	
Emancipação - Gramaticalização	

Fonte: Elaboração própria.

3.2.2 Teoria dos Exemplos

A Teoria dos Exemplos surge com o propósito de explicar o processo cognitivo que resulta nos fenômenos descritos pela Fonologia de Uso, como a redução da forma ou do significado. Assim, Pierrehumbert (2001) propõe enquadrar os princípios da Fonologia de Uso num modelo teórico oriundo da psicologia, a Teoria dos Exemplos,

A teoria dos exemplos foi introduzida pela primeira vez na psicologia como um modelo de percepção e categorização. Subsequentemente estendeu-se especificamente a sons de discurso por Johnson (1996) e Lacerda (no prelo), fornecendo um modelo altamente bem-sucedido da categorização de vogal em particular. Goldinger (1996) também aplica o modelo fortemente relacionado de Hintzman (1986) para modelar a identificação e reconhecimento de palavras.²⁵

Enquanto modelos gerativos clássicos lidam de forma satisfatória com casos de aplicação categórica ou não aplicação de regras em contextos específicos, a realidade da língua em uso revela muitos casos em que se tem uma regra variável, casos em que, mesmo diante de contexto para aplicação categórica de uma regra, podem haver restrições impostas pela frequência alta ou baixa de uma palavra ou até mesmo por fatores linguísticos como tonicidade,

²⁴ It appears that the productivity of a pattern, expressed in a schema, is largely, though not entirely, determined by its *type* frequency: the more items encompassed by a schema, the stronger it is, and the more available it is for application to new items. Thus, the English Past Tense -ed applies to thousands of verbs and is much more productive than any of the irregular patterns, which are highly restricted in the number of verbs to which they apply.

²⁵ Exemplar theory was first introduced in psychology as a model of perception and categorization. It was subsequently extended specifically to speech sounds by Johnson (1996) and Lacerda (in press), providing a highly successful model of vowel categorization in particular. Goldinger (1996) also applies the strongly related model of Hintzman (1986) to model the identification and recognition of words. (PIERREHUMBERT, 2001, p. 140)

ponto ou modo de articulação de um contexto precedente ou seguinte etc. Nesses casos, uma saída para muitos modelos pode ser a inclusão de um elevado número de regras lexicais para lidar com essas “exceções”; porém, isso não condiz com os princípios de economia descritiva tradicionalmente almejados por uma teoria linguística.

Esses casos não previstos por modelos gerativos clássicos, que se multiplicam em dados de uso vernacular, não podem mais ser ignorados, como se estivéssemos diante de exceções; já que, quantitativamente, essas exceções já chamam a atenção em muitos estudos sobre variação linguística e tornam, assim, necessária uma revisão de alguns paradigmas teóricos para poder acomodá-los.

É para responder a essa lacuna de modelos teóricos mais abstratos que surge a Teoria dos Exemplares. Esse modelo teórico propõe dar prosseguimento às consequências vislumbradas pela Fonologia de Uso, porém pretende ir além da descrição quantitativa e do controle estatístico de fatores linguísticos, problematizando a relação entre um modelo teórico categórico, como é o gerativo, e um modelo probabilístico (como por exemplo, o que propõe a Fonologia Estocástica): **qual é a melhor alternativa teórica para descrever os fenômenos não categóricos que ocorrem nas línguas do mundo?**

Em um modelo puramente gerativo tudo que não é categórico deveria pertencer à performance, ao uso, ou seja, não fazer parte essencialmente do conhecimento linguístico do falante. Já num modelo probabilístico, como a Teoria dos Exemplares, podemos ter uma combinação de regras abstratas e regras emergentes da frequência de *types* e *tokens*. Ou até mesmo, num recorte mais radical, apenas regras que emergem da frequência de uso de *types* e *tokens* (como é proposto por perspectivas funcionais e conexionistas da linguagem).

Pierrehumbert (2001, p.139) prefere a opção probabilística menos radical e admite a necessidade de combinar um modelo mais abstrato como o gerativo aos achados probabilísticos que uma fonologia baseada no uso busca descrever e explicar.

Embora uma palavra possa ter propriedades fonéticas idiossincráticas, ela é percebida como composta de unidades de estrutura sonora que também são compartilhadas com outras palavras. A existência dessas subpartes - sejam fonemas, moras ou sílabas - é refletida em comportamentos produtivos como a pronúncia de neologismos e assimilações de palavras emprestadas. Também se reflete na tendência de mudanças históricas para se espalhar pelo vocabulário. Assim, o modelo correto deve descrever a interação do detalhe fonético específico da palavra com princípios mais gerais da estrutura fonológica.²⁶

²⁶ Although a word may have idiosyncratic phonetic properties, it is perceived as made up of units of sound structure which are also shared with other words. The existence of these subparts—whether phonemes, moras, or syllables—is reflected in productive behaviors such as pronunciation of neologisms and loan word assimilations. It is also reflected in the tendency of historical changes to sweep through the vocabulary. Thus, the correct model

Pode-se talvez pensar em um modelo em que estão competindo permanentemente regras abstratas e regras emergentes do uso, sem que haja pressão da estrutura por uma ou outra rota de percepção e produção dos fenômenos linguísticos.

Para Pierrehumbert, a Teoria dos Exemplos pode oferecer uma arquitetura formal gerativa, porém inclui o detalhe fonético como parte da representação lexical, ou seja, como um elemento do conhecimento linguístico dos falantes.

[...] Essa arquitetura formal é "gerativa" no sentido de fornecer explicitamente para representações e processos fonológicos; prediz que alguns resultados são possíveis e outros não. Como uma gramática gerativa, ela é informada pelo objetivo de especificar todos e apenas os resultados possíveis na linguagem humana. Representa um afastamento considerável dos modelos gerativos, no entanto, na forma como as representações lexicais são organizadas e as consequências da representação lexical para a produção da fala (PIERREHUMBERT, 2001, p. 139).²⁷

Como consequência desse posicionamento teórico, a clássica distinção léxico/gramática, memória/processamento, palavra/regra não cabe mais num modelo que entende que

Memórias fonéticas detalhadas estão associadas a palavras individuais e, implicitamente, define distribuições de probabilidade específicas de palavras em relação a resultados fonéticos. Enquanto os modelos clássicos definem uma forte separação entre o léxico e a gramática, no presente modelo estes representam dois graus de generalização sobre as mesmas memórias e estão, portanto, fortemente relacionados entre si (PIERREHUMBERT, 2001, p. 139).²⁸

Assim a frequência é incluída como parte do conhecimento linguístico dos falantes, já que

Se cada token encontrado de uma categoria for armazenado como um exemplar separado, as categorias frequentes serão, obviamente, representadas por vários *tokens* e as categorias não frequentes serão representadas por *tokens* menos numerosos. A diferença na contagem de token é um ingrediente das explicações do modelo sobre os efeitos de frequência (PIERREHUMBERT, 2001, p. 140).²⁹

must describe the interaction of word-specific phonetic detail with more general principles of phonological structure.

²⁷ [...] This formal architecture is "generative" in the sense that it provides explicitly for phonological representations and processes; it predicts that some outcomes are possible and others are not. Like a generative grammar, it is informed by the goal of specifying all and only the outcomes which are possible in human language. It represents a considerable departure from generative models, however, in the way the lexical representations are organized and the consequences of lexical representation for speech production (PIERREHUMBERT, 2001, p. 139).

²⁸ Detailed phonetic memories are associated with individual words and it implicitly defines word specific probability distributions over phonetic outcomes. Whereas the classic models define a strong separation between the lexicon and the grammar, in the present model these represent two degrees of generalization over the same memories and are thus strongly related to each other (PIERREHUMBERT, 2001, p. 139).

²⁹ If every encountered token of a category is stored as a separate exemplar, then frequent categories will obviously be represented by numerous tokens and infrequent categories will be represented by less numerous tokens. The

Conforme Cristóforo Silva (2003, p. 211), a Teoria dos Exemplares oferece a noção de “nuvens de exemplares” para a representação de como se organizam os exemplares na memória dos falantes,

[...] Os exemplares são organizados num mapa cognitivo. Categorias mais frequentes apresentam maior número de exemplares e categorias infrequentes apresentam poucos exemplares. Uma nuvem de exemplares contém informação linguística e não-linguística. (CRISTÓFARO SILVA, 2003, p.11)

Quando falamos em detalhe fonético na representação lexical das palavras não podemos trivializar o papel da memória e suas limitações e, já que a frequência de *types* a que é exposto o falante interfere em sua representação mental desses mesmos *types*, Pierrehumbert comenta que “um indivíduo exemplar - que é uma memória perceptiva detalhada - não corresponde a uma única experiência perceptiva, mas sim a uma classe de equivalência de experiências perceptivas” (PIERREHUMBERT, 2001, p. 141).³⁰

Para concluir, segundo Pierrehumbert, a Teoria dos Exemplares oferece uma alternativa mais abrangente em relação aos modelos clássicos da fonologia gerativa porque toma como objeto de sua análise “o conhecimento fonético implícito dos falantes”.

A aquisição desse conhecimento pode ser entendida simplesmente em termos da aquisição de um grande número de traços de memória de experiências. Não existe um modelo competitivo que alcance o mesmo nível de adequação descritiva. Notavelmente, a suposição de que existe um alfabeto simbólico universal que fornece uma interface para um componente de implementação fonética sensorio-motora universal (como em Chomsky e Halle 1968; Chomsky e Lasnik 1995) não fornece nenhum meio de representar as diferenças extremamente sutis entre linguagens em valores e distribuições de probabilidade de propriedades fonéticas. Portanto, não fornece uma visão de como o conhecimento de tais detalhes pode ser adquirido. (PIERREHUMBERT, 2001, p. 143).³¹

Segundo Cristóforo Silva (2003, p. 213), podemos sintetizar a Teoria dos Exemplares em seis pontos:

(1) pretende expressar o conhecimento implícito que falantes possuem sobre os detalhes de dados fonéticos; (2) os sons são avaliados em contexto, sendo a palavra o

difference in token count is one ingredient of the model’s explanations of frequency effects (PIERREHUMBERT, 2001, p. 140).

³⁰ an individual exemplar—which is a detailed perceptual memory—does not correspond to a single perceptual experience, but rather to an equivalence class of perceptual experiences. (PIERREHUMBERT, 2001, p. 141).

³¹ The acquisition of this knowledge can be understood simply in terms of the acquisition of a large number of memory traces of experiences. There is no competing model which achieves the same level of descriptive adequacy. Notably, the assumption that there exists a universal symbolic alphabet which provides an interface to a universal sensori-motor phonetic implementation component (as in Chomsky and Halle 1968; Chomsky and Lasnik 1995) provides no means of representing the extremely fine differences across languages in values and probability distributions of phonetic properties. Therefore, it yields no insight into how the knowledge of such details might be acquired. (PIERREHUMBERT, 2001, p. 143).

locus da categorização; (3) procede a categorização com efeitos de protótipos; (4) a frequência é um mecanismo de diagnóstico do mapeamento fonológico; (5) permite incorporar resultados sócio-fonéticos à descrição; (6) implica que mudanças sonoras não são desfeitas (exemplares se tornam robustos ao aumentarem a frequência). (CRISTÓFARO SILVA, 20003, p. 213)

Desse modo, ao incluir a informação de detalhe fonético no modelo que descreve como se configura a linguagem na mente do falante, a Teoria dos Exemplares permite explicar o suposto “caos” do uso da linguagem, revelando a motivação linguística para os padrões que se observam em fenômenos que ocorrem no nível fonético, já que concretiza a ligação entre a regra abstrata dos modelos gerativos e os padrões de variação que emergem do uso pelas comunidades de fala.

3.3 O STATUS DA VOGAL TEMÁTICA NO PORTUGUÊS BRASILEIRO

Nesta seção apresentamos algumas propostas para a descrição da vogal temática nominal.

Os casos de derivação regressiva apresentam como vogal final *a* (*fuga*), *o* (*abandono*) e *e* (*combate*). Essas vogais são descritas tradicionalmente como pertencendo às três classes temáticas dos nomes (substantivos e adjetivos) no português brasileiro (PB) e são relacionadas à flexão do gênero masculino (maioria dos casos, palavras terminadas em *o*, e algumas vezes palavras terminadas em *e* ou *a*) ou feminino (maioria dos casos, palavras terminadas em *a*, e algumas vezes palavras terminadas em *e* ou *o*).

Compreender o *status* morfológico e fonológico desse segmento pode ajudar na descrição e explicação dos fatores que estão envolvidos na seleção das três vogais que emergem nas formas derivadas, já que nossa **hipótese** é de que **os padrões produtivos sinalizam que a derivação regressiva é um processo especial de nominalização na derivação regressiva que toma o radical como base da derivação, acrescentando uma vogal final que é influenciada por fatores linguísticos.**

3.3.1 Camara Jr. (1969,1970)

Para Camara Jr. (1969, p. 52), há um paralelo entre nome e verbos quanto à vogal temática, porque em ambas as categorias essa vogal distribui os vocábulos em classes morfológicas.

Para os verbos temos as três conjugações: primeira conjugação (*nadar*); segunda conjugação (*vencer*) e terceira conjugação (*partir*). No caso dos nomes, temos as classes temáticas *a* (*rosa, poeta*), *o/u* (*lobo, livro*) e *e/i* (*ponte, dente*). Além disso, há casos de vocábulos atemáticos, em que a vogal temática está ausente, como nomes terminados em consoante como *mar, sol* e aqueles terminados em vogal tônica como *café, alvará*. Para as formas terminadas em consoante, como *jantar*, a vogal temática pode emergir na flexão do plural, como em *jantares* (CAMARA JR., 1970, p. 85).

Conforme Camara Jr. (1970), a função da vogal temática é associar-se com o radical para formar o tema, em um processo análogo ao que se observa nas formas verbais. Esse processo não deve ser confundido com a flexão de gênero que forma feminino em *a* e, masculino na ausência do sufixo de gênero *a*.

Há nos nomes os temas em *a* (*rosa, poeta, planeta*), os temas em *o /u/* átono final (*livro, tribo, cataclismo*) e os temas em *e /i/* átono final (*dente, ponte, análise*). Assim não se confunde a desinência de feminino *a*, que aparece especialmente nos temas em *o* (*lobo, loba*) e a vogal temática em *a*, que não é marca de gênero (ci. *poeta*, masculino; *artista*, masculino ou feminino conforme o contexto). (CAMARA JR., 1970, p. 85).

3.3.2 Kehdi (1999)

Alinhando-se à descrição mattosiana, Kehdi (1999) destaca que as formas nominais terminadas em vogal tônica, considerados atemáticas, como *café, alvará, saci, camelô* etc., não sofrem apagamento da vogal final quando servem de base para a derivação sufixal. Segundo o autor, quando estamos verdadeiramente diante de uma vogal temática, ocorre o seu apagamento em casos de derivação sufixal (*casa > caseiro*)

Assim, se temos uma forma terminada em vogal temática como *bolo*, o acréscimo do sufixo *-inho* forma *bolinho* e não **boloinho*, ou seja, a vogal temática é apagada na derivação; já diante de *café*, temos *cafezinho* e não **cafinho*; o que parece demonstrar que essa vogal tônica não pode ser confundida com uma vogal temática nominal.

3.3.3 Lee (1995)

Lee (1995), inspirado nas ideias de Harris (1991) sobre vogal temática, propõe a noção de classes temáticas como marcadores de palavras. Esses marcadores teriam como função principal emergir para “completar a palavra” (LEE, 1995, p. 34). Sem esse marcador, as

palavras não podem emergir, já que apresentariam formas agramaticais (**menin*, **estudant*, **poet*, por exemplo).

Com base na arquitetura da Morfologia e Fonologia Lexical, com um léxico descrito em ciclos de flexão e derivação, Lee (1995, p. 40) traz casos de vocábulos como *matagal* e *chubarada*, que comprovam que a vogal temática deve preceder o ciclo da flexão, já que ela pode ocorrer entre constituintes: [[mat[**a**]][gal]]; [[chuv[**a**]][rada]]. Essa perspectiva contraria a proposta de Harris (1991), que posiciona esses marcadores no nível flexional (como o -s de plural).

Para Lee, a vogal temática tem uma função morfológica, atribuir o gênero, e uma função fonológica, dar condições para que as palavras respeitem as restrições fonotáticas para a formação de vocábulos em PB (*Condição de Perifericidade*, conforme Inkelas, 1989, Hayes, 1991). Esse estatuto fonológico permitido pela *Condição de Perifericidade* pode ser perdido se for acrescido um sufixo derivacional, como na palavra derivada *caseiro*, em que ocorre o apagamento do marcador -a.

3.3.4 Moreno (1997)

Também com base na análise de Harris (1991) para as vogais finais do espanhol, Moreno (1997) considera as vogais temáticas como marcadores ou elementos terminais. Para o autor a proposta para a língua espanhola pode enquadrar o PB sem alterações: classe I, palavras terminadas em -o; classe II palavras terminadas em -a, classe III, palavras terminadas em -e, sem distinção para os casos em que a vogal é ou não epentética. Moreno aceita a ideia de que os marcadores só podem ocorrer no nível flexional e nunca no nível derivacional, o que se opõe ao que é apontado por Lee (1995), nos casos em que mesmo após a derivação a vogal temática emerge, como em *matagal*, ou seja, não é apagada como em casos como *caseiro* ou *lojista*. Ainda em consonância com o que é proposto por Harris, Moreno retoma a noção de que a vogal temática é um morfema flutuante, “um segmento que não está ligado ao esqueleto prosódico nas representações fonológicas iniciais. Num momento posterior, na derivação, os marcadores possíveis são incorporados à estrutura prosódica e só então recebem uma interpretação fonética” (MORENO, 1997, p. 48-49).

Moreno (1997, p. 57) classifica os marcadores de palavras “em três classes hierárquicas: um **núcleo central** de protótipos; um **núcleo periférico** de casos levemente desviantes, especialmente os vocábulos que não têm marcador, e um variado **resíduo**”.

O **núcleo central** se refere aos casos mais produtivos, em que há correlação entre marcador e gênero. **Núcleo periférico** são as formas terminadas em -e ou consoantes, em que a correlação marcador e gênero se torna imprevisível. E **resíduo** se refere aos vocábulos em que a correlação marcador e gênero é inversa ao esperado em PB, como *o problema* e *a tribo*.

3.3.4 Alcântara (2010)

Alcântara (2010) propõe a existência de quatro classes formais para os nomes do PB, incluindo, assim, mais uma às tradicionais três classes temáticas. Para a autora, as vogais *o*, *a* e *e* são morfemas do português (a vogal *e* também pode ser uma vogal epentética).

Quadro 2- Classes formais do português

Classe Formal	
I. palavras terminadas em <i>o</i>	masculinas: astro, belo, dado, figo femininas: libido, tribo, virago
II. palavras terminadas em <i>a</i>	masculinas: aroma, cometa, drama femininas: bela, dama, cama
III. palavras terminadas em <i>e</i>	masculinas: abacate, açougue, alarde femininas: arte, ave, boate masc./femin.: célere, mole, precoce
IV. palavras sem vogal final	masculinas: frei, rei, boi femininas: bagagem, coragem, pá masc./femin.: jovem

Fonte: Adaptado de ALCÂNTARA, 2010, p. 6.

A numeração das classes propostas pela autora indica o ranqueamento de frequência das terminações, da classe mais frequente *o* (I), passando por *a* (II), *e* (III) até a classe menos produtiva (IV) que é composta de formas atemáticas, como *coragem* e *lei*.

Em suma, todas as raízes masculinas não-marcadas para classe, assim como as femininas portadoras do traço de classe formal [I], recebem /o/, como manifestação fonológica/subjacente do morfema de classe formal representativo da Classe I, a classe *default* [...]. Por ser considerada *default*, tal classe é consequentemente a mais produtiva da língua, ao lado da classe formal II. (ALCÂNTARA, 2010, p. 10)

Esse ordenamento não é condizente com o observado por Schwindt (2018), que aponta que, no português brasileiro, as formas terminadas em *a* são mais frequentes que formas terminadas em *o*. A divergência neste ponto se dá devido a uma diferença metodológica: ao contrário de Schwindt, Alcântara não oferece uma base quantitativa para sustentar seu ranqueamento entre os nomes terminados em *o* e *a*. Conforme a própria autora,

Os dados a serem apresentados não sofreram tratamento quantitativo, em virtude do caráter estritamente teórico do trabalho. Salienta-se, não obstante, que as fontes consultadas foram inúmeras – a partir delas construíram-se as listas não exaustivas de vocábulos nominais do português, subsequentemente organizadas em diferentes classes formais, seguindo a proposta de Harris (1999). (ALCÂNTARA, 2010, p.5)

Assim, o pertencimento de nomes terminados em *o* na Classe I reflete apenas a percepção da autora da frequência superior de nomes com essa vogal final no português brasileiro, “A Classe I, identificada pelo morfema de classe formal /o/, parece ser a que mais palavras abriga na língua, sendo, por isso, um dos agrupamentos formais mais produtivos” (ALCÂNTARA, 2010, p. 6).

Assim como já referido por Camara Jr. (1969, 1970), Alcântara menciona que não se devem confundir as vogais temáticas, que separam formas nominais em classes formais, das desinências de gênero *o* e *a*, já que as classes I e II reúnem formas masculinas e femininas (mesmo que a predominância de formas terminadas em *o* masculinas e de formas terminadas em *a* femininas possa provocar essa percepção de identidade entre vogal temática e desinência de gênero³²).

Com base nos princípios da Morfologia Distribuída, Alcântara (2010) descreve da seguinte forma as regras que envolvem a emergência das três classes formais nas formas nominais do PB³³:

1. **Classe I** (*o*): como a maioria das formas dessa classe apresenta o **gênero masculino**, o qual é não marcado no português (*default*), as entradas lexicais dessa classe não carregam informação sobre a classe a qual estão associadas. Somente carregam informação de classe quando são femininas (*libido, tribo*).

³² Ver Schwindt (2018) para uma discussão mais aprofundada sobre a existência de subregularidades que justificam essa impressão por parte dos falantes de PB.

³³ A autora não trata dos nomes da Classe IV, cujo morfema de classe formal é zero fonológico.

Figura 2 - Membros da Classe I.

a. <i>entradas vocabulares das raízes</i>		
/libid/, f, I		
/menin/		
b. <i>derivações</i>		
<i>libido</i>	<i>menino</i>	
[/libid/]ǂ	[/menin/]ǂ	1 MORFOLOGIA
[]	[]	
f		a
I		b
↓	↓	2 <i>Inserção vocabular</i>
o	o	
libid+ <u>o</u>	menin+ <u>o</u>	

Fonte: ALCANTARA, 2010, p. 9.

2. **Classe II (a)**: como a maioria das formas dessa classe apresenta o **gênero feminino**, o qual é marcado no português, as entradas lexicais dessa classe carregam informação sobre o gênero feminino a qual estão associadas. Somente carregam informação de classe quando são masculinas (*tema, problema* etc.).

Figura 3 - Membros da Classe II

a. <i>entradas vocabulares das raízes</i>		
/komet/, II		
/mal/, f		
b. <i>derivações</i>		
<i>cometa</i>	<i>mala</i>	
[/komet/]ǂ	[/mal/]ǂ	1 MORFOLOGIA
[]	[]	
	f	a
II		b
	II	2
↓	↓	3 <i>Inserção vocabular</i>
a	a	
komet+ <u>a</u>	mal+ <u>a</u>	

Fonte: ALCANTARA, 2010, p. 10.

3. **Classe III (e)**: as formas incluídas nessa classe podem se referir a palavras terminadas em *e* como *pente, gente*, como palavras terminadas em consoantes que fazem emergir a vogal temática quando são flexionadas no plural, como em *par > pares*. Se emergem como palavras masculinas, carregam informação de classe apenas, a fim de não serem incluídos na classe

default (I); já se resultam em vocábulos femininos, precisam carregar informação sobre o gênero e sobre a classe, para não emergirem como membros da classe II.

Figura 4 - Membros da Classe III

(11) Classe III: /e/ ~ Ø

<p>a^o. /S/ algoS-Ø/algoz-<u>es</u> cruS-Ø/cruz-<u>es</u> feliS-Ø/feliz-<u>es</u> /r/ ar-Ø/ar-<u>es</u> dever-Ø/dever-<u>es</u> flor-Ø/flor-<u>es</u> mártir-Ø/mártir-<u>es</u></p> <p>a^o. /s/ alfa/s/-<u>e(s)</u> cla/s/-<u>e(s)</u> mu/s/-<u>e(s)</u> /r/ folclor-<u>e(s)</u> escor-<u>e(s)</u> árvor-<u>e(s)</u> /l/ control-<u>e(s)</u> mol-<u>e(s)</u> pel-<u>e(s)</u> val-<u>e(s)</u></p>	<p>b. C av-<u>e(s)</u> blef-<u>e(s)</u> cra/z/-<u>e(s)</u> club-<u>e(s)</u> cra/k/-<u>e(s)</u> deta/l/-<u>e(s)</u> debo/f/-<u>e(s)</u> pirâmid-<u>e(s)</u> pre/jn-<u>e(s)</u> tim-<u>e(s)</u> sorvet-<u>e(s)</u></p> <p>CC alarm-<u>e(s)</u> alegr-<u>e(s)</u> bail-<u>e(s)</u> bandeid-<u>e(s)</u> bos/k/-<u>e(s)</u> carn-<u>e(s)</u> célebr-<u>e(s)</u> ciprest-<u>e(s)</u> chan/s/-<u>e(s)</u> charm-<u>e(s)</u> cism-<u>e(s)</u> client-<u>e(s)</u> consort-<u>e(s)</u> entors-<u>e(s)</u> film-<u>e(s)</u> humild-<u>e(s)</u> par/k/-<u>e(s)</u> urz-<u>e(s)</u></p>
---	---

Fonte: ALCÂNTARA, 2010, p. 11.

3.3.5 Schwindt (2018)

Schwindt (2018), ao abordar a relação entre gênero gramatical e classe temática a partir de dados de produtividade do português brasileiro, sob a perspectiva da HS-OI³⁴ ((McCarthy 2011), propõe que “A implementação dessas formas se define na base da competição entre

³⁴ Modelo que combina o Serialismo Harmônico (McCarthy 2007) com uma abordagem da Teoria da Otimidade (OT) conhecida como Optimal Interleaving (Wolf 2008).

restrições de fidelidade, responsáveis pela realização de morfemas, e restrições de marcação fonológica, que dão conta da boa formação da sílaba e da palavra prosódica” (SCHWINDT, 2018, p. 746).

Para sustentar a desvinculação entre realização morfológica e fonológica dos morfemas de gênero no português brasileiro, o autor (SCHWINDT, 2018, p. 751) aponta como exemplo o fato de numa frase como “o linguista atento” a concordância com o gênero masculino do artigo definido “o” e do adjetivo “atento” ocorrer apesar da vogal final *a* da palavra linguista, típica de formações de nomes de gênero feminino, o que parece indicar a existência de um alomorfe zero para esse gênero gramatical.

Nossa hipótese, então, é a de que morfemas, que nascem abstratos, estão sujeitos a exponenciação (cf.: Mathews 1974), isto é, são realizados fonologicamente, como preconizado, por exemplo, pela DM ou pela HS-OI. No percurso da exponenciação (ou do mapeamento entre morfologia e fonologia), operam restrições de fidelidade e de marcação fonológica. (SCHWINDT, 2018, p. 751)

Com base em consulta a léxico dicionarizado (Dicionário Eletrônico Aurélio), uso vernacular (Projeto VARSUL) e corpus de referência (Projeto ASPA), o autor conclui que

Há, no mapeamento geral da língua, isomorfismo entre a terminação *o* e gênero masculino, de um lado, e entre a terminação *a* e *o* gênero feminino, de outro, com alguma vantagem de frequência de feminino. Essa correlação diz respeito, majoritariamente, a gênero gramatical, não sexo biológico ou gênero social. Se observados, contudo, subconjuntos de dados, como o dos poucos nomes que estabelecem pareamento com sexo biológico ou gênero social, e mesmo o dos biformes e comuns de dois na língua em uso, apesar de sua frequência também limitada, constata-se que o isomorfismo mencionado se mantém, e as formas masculinas passam a prevalecer. (SCHWINDT, 2018, p. 766)

Essas conclusões indicam essencialmente a frequência superior de substantivos femininos, o que contraria o senso comum e algumas propostas teóricas (como proposto por ALCÂNTARA, 2010) que apontam o masculino como a forma nominal mais frequente no português brasileiro. Essa predominância do masculino parece restrita ao subconjunto de formas que fazem referência ao gênero biológico ou social, conforme apontado acima pelo autor, o que pode estar conduzindo à percepção de predomínio de formas masculinas por parte dos falantes do português brasileiro. É importante salientar que Schwindt considerou em sua análise apenas substantivos que não eram “compostos, derivados de outros nomes e elementos de locuções” (SCHWINDT, 2018, p. 752).

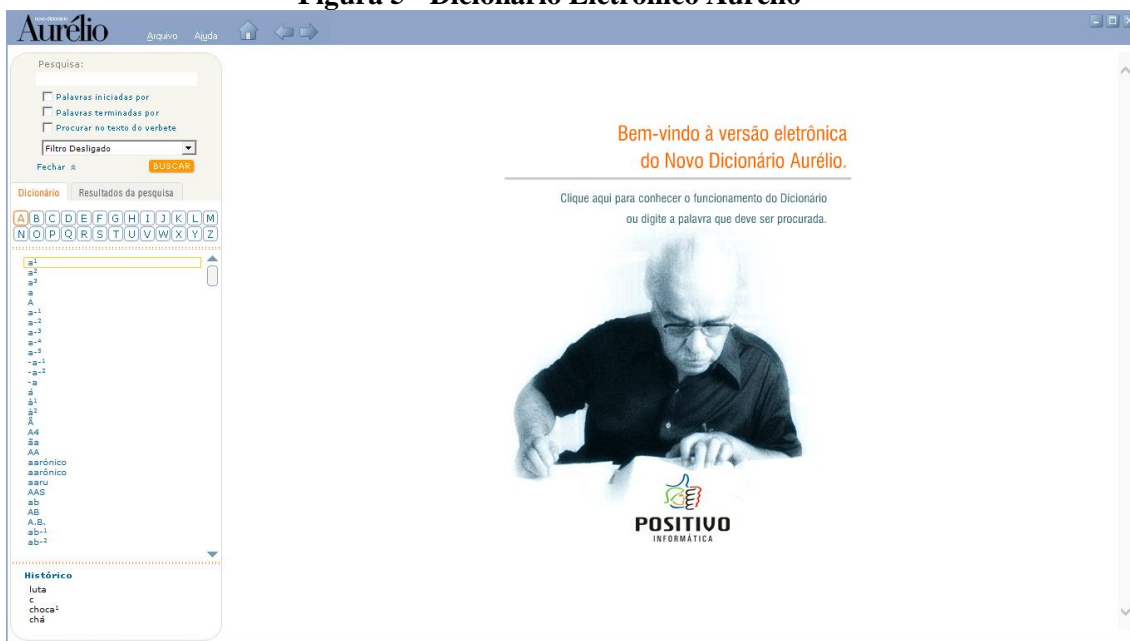
4 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo apresentamos as etapas que constituíram nossos procedimentos metodológicos: (i) o levantamento de dados no léxico dicionarizado, por meio de consulta aos dicionários eletrônicos *Aurélio* e *Houaiss*, (ii) a consulta ao léxico institucionalizado, para registro da frequência lexical dos dados no *Corpus Brasileiro*, (iii) a codificação segundo fatores linguísticos, (iv) comparação entre a frequência de *tokens* e *types*, (v) levantamento dos padrões sonoros mais frequentes no final da palavra, (vi) análise do léxico potencial, por meio da aplicação de um experimento online a 100 informantes e (vii) a comparação entre resultados do levantamento de dados do léxico dicionarizado, do léxico institucionalizado e do léxico potencial.

4.1 LÉXICO DICIONARIZADO

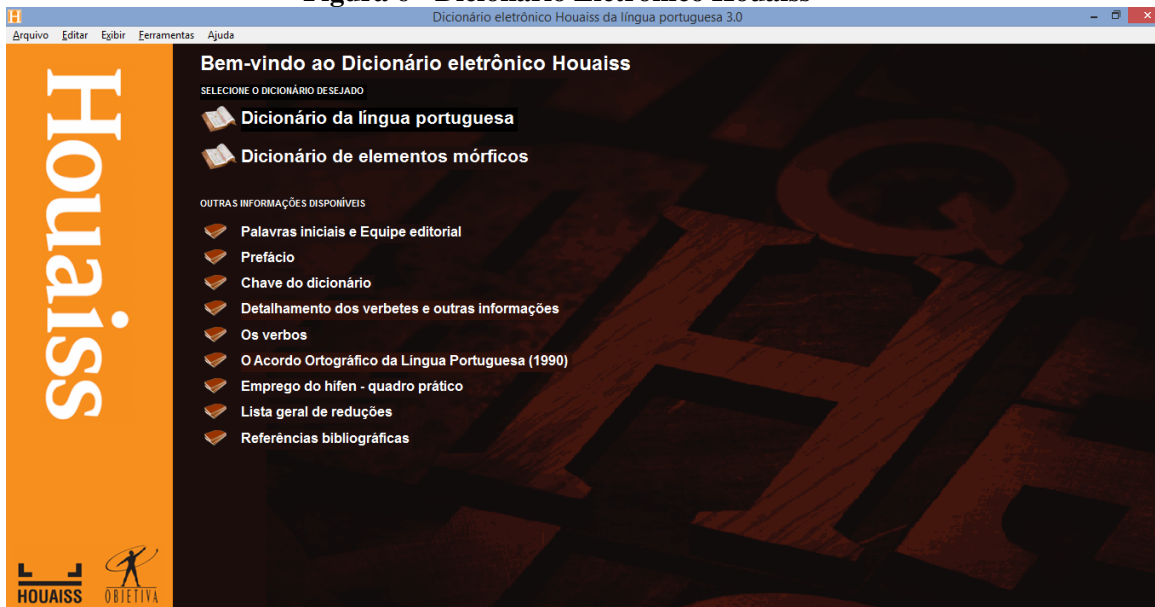
Na Etapa 1, procedemos ao levantamento de nomes deverbais registrados nos dicionários eletrônicos *Aurélio* (versão 5.0.40) e *Houaiss* (versão 3).

Figura 5 - Dicionário Eletrônico Aurélio



Fonte: Dicionário Eletrônico Aurélio (versão 5.0.40).

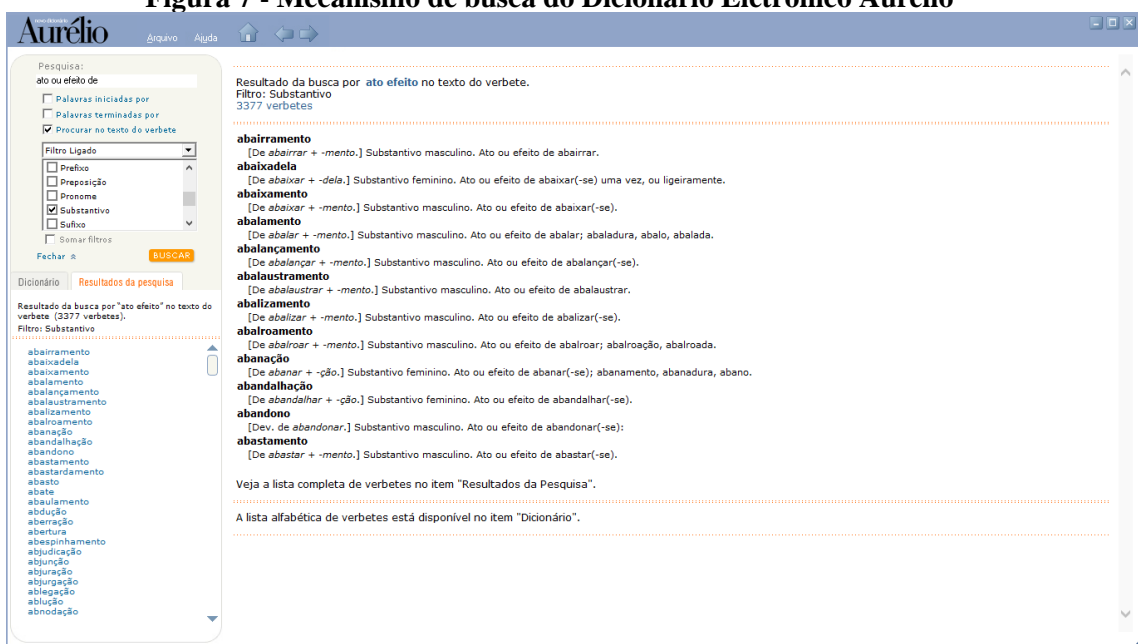
Figura 6 - Dicionário Eletrônico Houaiss



Fonte: Dicionário Eletrônico Houaiss (versão 3)

Foram utilizados os filtros “Procurar no texto do verbete” e “Substantivo” (recursos oferecidos pelo dicionário Aurélio para restringir a busca de palavras). Assim na caixa de texto digitamos “ato ou efeito de” (trecho que aparece nos substantivos derivados de verbos). Essas palavras foram analisadas individualmente para selecionar apenas os casos de substantivos deverbais via derivação regressiva, ou seja, apenas os casos definidos pelo dicionário como nomes deverbais.

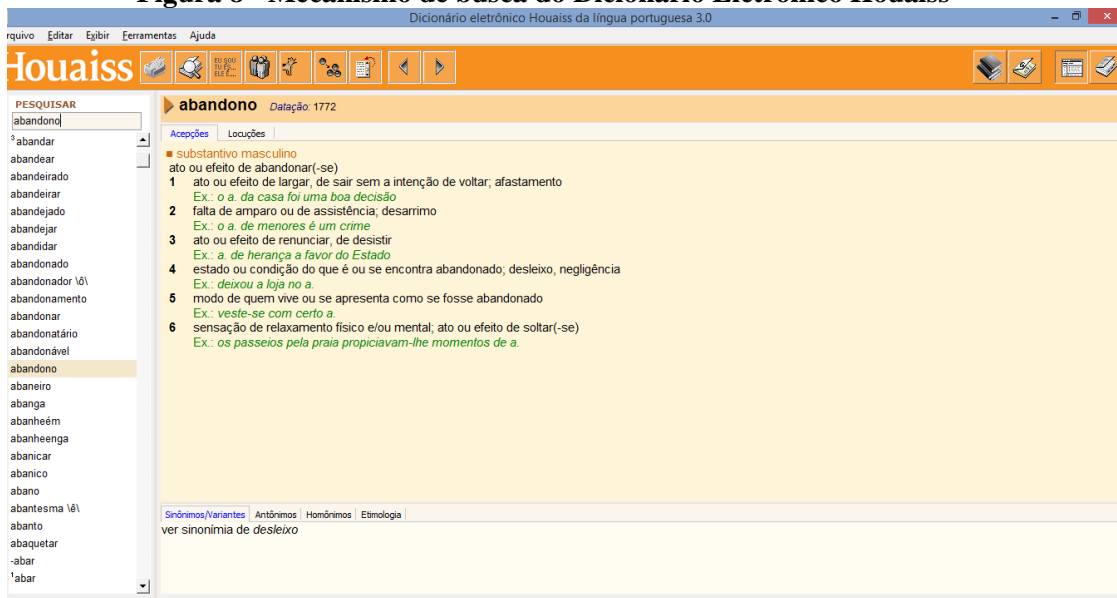
Figura 7 - Mecanismo de busca do Dicionário Eletrônico Aurélio



Fonte: Dicionário Eletrônico Aurélio (versão 5.0.40).

O Dicionário Houaiss foi utilizado para corroborar a existência das palavras encontradas no Aurélio, e também para anotar os casos em que havia informação sobre a datação do primeiro registro escrito dos nomes deverbais analisados.

Figura 8 - Mecanismo de busca do Dicionário Eletrônico Houaiss



Fonte: Dicionário Eletrônico Aurélio (versão 5.0.40).

4.2 LÉXICO INSTITUCIONALIZADO

Em relação à escolha de qual *corpus* serviria de base para nossa análise da derivação regressiva no contexto de uso dos falantes do português brasileiro, seguimos os critérios estabelecidos por Sardinha (2000). Segundo o autor (SARDINHA, 2000, p. 338), para que um conjunto de dados linguísticos possa representar um recorte sincrônico representativo de uma língua natural devem ser levados em consideração os seguintes critérios:

- (a) A origem: Os dados devem ser autênticos.
- (b) O propósito: O *corpus* deve ter a finalidade de ser um objeto de estudo linguístico.
- (c) A composição: O conteúdo do *corpus* deve ser criteriosamente escolhido.
- (d) A formatação: Os dados do *corpus* devem ser legíveis por computador.
- (e) A representatividade: O *corpus* deve ser representativo de uma língua ou variedade.
- (f) A extensão: O *corpus* deve ser vasto para ser representativo. (SARDINHA, 2000, p. 338)

Dessa forma, na Etapa 2, baseamo-nos no conjunto de dados linguísticos do Corpus Brasileiro, que é um banco de textos falados e escritos com 938.7³⁵ milhões de palavras registradas. Ali buscamos a frequência de ocorrência de cada uma das palavras levantadas no Dicionário Aurélio e corroborada no Dicionário Houaiss. Por meio do site Linguateca³⁶, com recursos que facilitam o acesso ao Corpus Brasileiro, utilizamos a ferramenta de busca “Procurar por Distribuição da categoria gramatical (PoS)”, que nos indicava a frequência de substantivos para a forma procurada, já que era essencial distinguir os casos em que há coincidência da forma verbal flexionada e da forma nominal derivada (por exemplo, “abandono”, que pode se referir à primeira pessoal do singular do presente do indicativo do verbo abandonar, ou ao nome derivado do verbo abandonar).

Figura 9 - Mecanismo de busca do Corpus Brasileiro no site linguatca.pt
Projeto AC/DC: corpo Corpus Brasileiro

AC/DC : Linguatca

O [Corpus Brasileiro](#) é uma coletânea de aproximadamente um bilhão de palavras de português brasileiro, resultado de projeto coordenado por [Tony Berber Sardinha](#) (GELC, LAEL, Cepril, PUCSP), com financiamento da [Fapesp](#). Devido a variados tipos de processamento automático, é possível que nem todo o material incorporado esteja disponível através do AC/DC.

Procurar:

Resultado:

- Concordância
- Distribuição das formas (*word*)
- Distribuição dos lemas (*lema*)
- Distribuição da categoria gramatical (PoS) (*pos*)
- Distribuição do tempo verbal e/ou do caso pronominal (*temcase*)
- Distribuição de pessoa e/ou número (*personum*)
- Distribuição do gênero morfológico (*gen*)
- Distribuição da função sintática (*func*)
- Distribuição por gênero de texto (*genero*)
- Distribuição por campo semântico (*sema*)
- Distribuição por grupo (de cor, roupa, etc.) (*grupo*)

Opções

- Resultados por ordem alfabética (só distribuições)
- Resultados em formato separado por ponto e vírgula
- Ignorar maiúsculas/minúsculas (não admite parâmetros)

Fazer nuvem com limite de

Amostra aleatória de linhas.

Tipo	Diversos
Variante(s)	BR
Tamanho (unidades)	1110.4 milhões
Tamanho (palavras)	939.5 milhões

Carateres úteis: | { } []

[Página principal](#)

Procure noutros corpos:

[AmostrA-NILC ANCB Avante! Corpus Brasileiro CD HAREM CETEMPúblico CHAVE Ciência Viva Coloma CONDIVport CONDIVport2 CoNE C-Oral-Brasil DHBB DiaCLAV Diáspora TL-PT ECL-EBR ECL-EE ENPCPUB \(parte em português\) Floresta FrasesPB FrasesPP Mariano Gago Literatca Marielle, presente! Moçambula Museu da Pessoa Natura/Minho NOBRE OBRas Plo Norte Português Falado - Documentos Autênticos ReLi NILC/São Carlos todos juntos Tycho Brahe Vercial](#)

Fonte: Site linguatca.pt.

Figura 10 - Informações de frequência no Corpus Brasileiro
Resultados da procura

20 de janeiro de 2020

Procura: "abandono"
 Distribuição de pos
 Corpo: Corpus Brasileiro v: 5.1

17238 casos.

Distribuição

Houve 6 valores diferentes de pos.

N	16677
abandono	430
V	125
abandonar	2
V-REGENTE-NEG	2
PROP_tit	2

Para informação sobre os códigos internos do atributo pos, consulte a página de [anotação](#).

[voltar]
 [nova pesquisa]

[Perguntas, comentários e sugestões](#)

Fonte: Site linguatca.pt.

³⁵ Por ser um banco ao qual estão sempre sendo acrescentadas novos textos, o número de palavras registradas e suas respectivas frequências estão sempre sendo atualizadas.

³⁶ <https://www.linguatca.pt/acesso/corpus.php?corpus=CBRAS>

4.3 CODIFICAÇÃO SEGUNDO FATORES LINGUÍSTICOS

Na Etapa 3, os nomes deverbais levantados no dicionário Aurélio, corroborados no dicionário Houaiss e com frequência de no mínimo uma ocorrência no Corpus Brasileiro foram codificados (códigos listados no Apêndice D) segundo as seguintes variáveis linguísticas, apresentadas sob a forma de grupos de fatores.

4.3.1 Vogal temática verbal

Buscamos responder a seguinte questão: há uma relação entre a vogal temática dos verbos derivantes e a vogal temática dos nomes derivados?

Por exemplo, pode-se afirmar que, diante de uma palavra como *atacar*, há uma seleção da vogal temática do nome deverbal *ataque* (vogal final *e*) com base na informação sobre a vogal temática verbal (vogal temática *a-*)? Ou seja, pode-se defender a hipótese de que cada uma das três vogais temáticas restringem as possibilidades de expressão das três vogais temáticas nominais.

- (i) **primeira conjugação:** canto (v. cantar)
- (ii) **segunda conjugação:** perda (v. perder)
- (iii) **terceira conjugação:** consumo (v. consumir)

Acreditamos, com base na Teoria dos Exemplos, que os verbos com vogal temática *a* por serem os mais produtivos selecionam com uma maior frequência palavras com vogal temática nominal *o*, dado que é aparentemente a vogal temática mais frequente (BASILIO, 2011), e conseqüentemente a esse par é associado o gênero masculino, já que em geral é atribuído esse gênero à vogal temática *o*.

4.3.2 Identidade entre base verbal e nome derivado

Buscamos responder a seguinte questão: o que é mais frequente, a alternância da vogal temática da base verbal, como em atacAr/ataquE, ou a preservação da vogal, como em falAr/falA?

- (i) **preserva a vogal temática da base:** dança (v. dançar)
- (ii) **não preserva a vogal temática da base:** abandono (v. abandonar)

A partir dos pressupostos exemplaristas, defendemos a hipótese de que um número muito baixo de casos de preservação pode corroborar a proposta de uma base sem vogal temática (por exemplo, *danç-*, em *dançar*), já que o falante parece não ter acesso a essa informação quando produz formas de derivação regressiva (LOBATO, 1995).

4.3.3 Complexidade morfológica

Buscamos responder a seguinte questão: qual é o tipo preferido de base para emergência da vogal final: palavras simples, palavras prefixadas ou palavras sufixadas?

(i) **palavra simples:** abraço

(ii) **palavra complexa:** desajuste

Tomamos como hipótese a questão de as palavras sufixadas, por já terem passado por um processo de adição de um segmento à direita da base, apresentarem um número baixo de casos de derivação regressiva; por outro lado, formas simples e prefixadas podem apresentar um número maior de casos desse tipo de derivação, já que não tiveram nenhum sufixo afixado (conforme RESENDE, 2016, podemos estar diante de sufixação zero).

4.3.4 Vogal média tônica

Buscamos responder a seguinte questão: formas verbais que apresentam alternância vocálica nas formas rizotônicas podem ser consideradas um indício de que as formas regressivas não tomam como base a forma flexionada, mas sim o morfema de radical como base para a derivação?

(i) **presença de vogal média tônica:** desfolha

(ii) **ausência de vogal média tônica:** derrame

Defende-se aqui a hipótese de que nomes deverbais preservem a vogal média da base verbal.

4.3.5 Gênero da palavra

Buscamos responder a seguinte questão: os nomes formados via derivação regressiva seguem a tendência dos nomes em geral em PB quanto ao gênero mais frequente?

(i) **gênero masculino:(o)** acordo

(ii) gênero feminino: (a) entrega

Há número mais frequente de palavras formadas via derivação regressiva do gênero masculino do que do gênero feminino, em outras palavras, um número maior de palavras terminadas em *o*, seguidas de *e* e *a* (BASILIO, 2011). Por outro lado, conforme Schwindt (2018), há mais palavras femininas do que masculinas no léxico do português brasileiro, o que torna mais interessante ainda a aplicação de um experimento online que possa verificar qual das duas forças tem mais relevância para a emergência da vogal final: padrão feminino do léxico ou padrão de formação da derivação regressiva?

Tomamos como hipótese, com base nos pressupostos da Teoria dos Exemplos, a relevância maior do padrão de formação da derivação regressiva para a emergência da vogal final.

4.3.6 Número de sílabas

Buscamos responder a seguinte questão: palavras com duas ou três sílabas são um alvo mais frequente para a emergência da vogal final? Ou o número de sílabas não é um fator relevante para a emergência da vogal final na derivação regressiva?

(i) dissílabos: com.pra

(ii) trissílabos: a.bri.go

(iii) polissílabos: ar.re.ma.te

Defendemos a seguinte hipótese: os nomes deverbais seguem a tendência dos nomes em geral no português brasileiro quanto ao número de sílabas, já que não consideramos este um fator relevante para a emergência da vogal final na derivação regressiva.

4.3.7 Onset da última sílaba

Buscamos responder a seguinte questão: há influência do segmento que precede a vogal final dos nomes deverbais? Será que ter um *onset* favorece o processo de derivação?

(i) presença de *onset*: perda

(ii) ausência de *onset*: passeio

Tomamos como hipótese a influência da presença ou não do *onset* para a emergência da vogal final: a ausência de *onset* influencia a emergência da vogal final *o*.

4.3.8 Modo de articulação do contexto fonológico precedente

Buscamos responder a seguinte questão: a frequência de determinado modo de articulação do segmento precedente pode estar condicionando a emergência da vogal final?

(i) **nasal:** acannho

(ii) **plosiva:** corte

(iii) **fricativa:** ceya

(iv) **lateral:** acúmulo

(v) **tepe:** conjra

(vi) **vogal:** bloqueio

Com base na Teoria dos Exemplos, partimos da hipótese de que alguns padrões sequenciais com frequência mais elevada podem ter um papel relevante para fixar na mente dos falantes a associação entre contexto fonológico precedente e vogal final em nomes deverbais.

4.3.9 Ponto de articulação do contexto fonológico precedente

Buscamos responder a seguinte questão: A frequência de determinado ponto de articulação do contexto fonológico precedente pode estar condicionando a emergência da vogal final?

(i) **bilabial:** derrube

(ii) **labio-dental:** desoya

(iii) **alveolar:** desajuste

(iv) **palatal:** desencaixe

(v) **velar:** pesca

(vi) **vogal alta:** balbucio

(vii) **vogal média:** amontoga

Com base na Teoria dos Exemplos, partimos da hipótese de que alguns padrões sequenciais com frequência mais elevada podem estar tendo um papel relevante para fixar na mente dos falantes a associação entre contexto fonológico precedente e vogal final em nomes deverbais.

4.4 COMPARAÇÃO ENTRE FREQUÊNCIA DE *TYPES* E *TOKENS*

Na Etapa 4, comparamos os resultados das variáveis linguísticas em função da frequência de *types* (conforme as palavras levantadas pelo dicionário) e da frequência de *tokens* (conforme a frequência de ocorrência de cada palavra no Corpus Brasileiro).

4.5 PADRÃO DE SEQUÊNCIA SONORA

Na etapa 5, fizemos um levantamento manual dos padrões de sequências sonoras finais mais comuns entre os nomes deverbais listados, comparando frequência de *types* e *tokens*.

4.6 LÉXICO POTENCIAL: EXPERIMENTO ONLINE COM PSEUDOPALAVRAS

Em nossos procedimentos metodológicos, percebemos que há uma limitação bastante importante em nossa análise. Como tomamos como ponto de partida o léxico dicionarizado para compor nossa lista de *types*, estamos, devido a essa característica metodológica, limitados a registrar a frequência lexical dessas palavras apenas, não sendo capazes de buscar formas não previstas pelos dicionários utilizados em nossa análise. Por exemplo, a palavra *desmate* não consta nos dois dicionários que usamos em nossa pesquisa; no entanto, a manchete de uma matéria do jornal Folha de São Paulo, de 16 de agosto de 2019, apresenta essa forma não registrada na seguinte frase: “ONG aponta alta de 15% na taxa de desmate da Amazônia e confirma tendência”. Esse exemplo ilustra uma grande dificuldade de nossa análise no que se refere a mapear formas novas de derivação regressiva que emergem nos padrões de uso dos usuários do PB: palavras que não estão registradas no dicionário mas que estão no *corpus* não apareceram. Para compensar essa limitação metodológica, elaboramos um experimento que contempla o aspecto potencial do fenômeno da derivação regressiva, registrando os possíveis padrões de produtividade não abarcados pela análise do léxico dicionarizado e do léxico institucionalizado.

Desse modo, na etapa 6, aplicamos um experimento com pseudopalavras (palavras potenciais no português brasileiro), que foram elaboradas a partir da frequência observada para os padrões fonológicos, morfológicos e de sequência sonora final de nomes deverbais via derivação regressiva. Esse tipo de experimento é essencial para depreendermos o conhecimento

linguístico internalizado dos falantes, ou seja, que tipo de regra está sendo acionada pelos falantes para a emergência da vogal final nas formações por derivação regressiva.

A partir dos resultados observados no léxico dicionarizado e no léxico institucionalizado, foi construído um experimento online (Apêndice C), por meio da plataforma digital Google Forms, com a participação de 100 informantes matriculados em cursos de graduação³⁷ no território brasileiro.

O experimento é constituído de um questionário com seis perguntas de múltipla escolha, utilizando-se pseudoverbos na forma infinitiva (*maluter, *fetadir, *zenecar) para que os informantes escolham uma das três opções de forma derivada (com vogal final *a*, *e* ou *o*), e pseudoverbos na forma derivada (*mopito, *beplida, *fituque), para que os informantes selecionem a forma infinitiva correspondente a partir de três opções possíveis (as três conjugações verbais, *-ar*, *-er* ou *-ir*). As pseudopalavras utilizadas no experimento baseiam-se nos tipos de terminações mais frequentes observados em levantamento realizado com base nos dicionários eletrônicos Houaiss e Aurélio. Por exemplo, em relação às três questões com pseudoverbos na forma infinitiva, diante da frase ” João costuma MALUTER diariamente. Ele, contudo, precisa comer muito antes de”, os informantes tinham quatro opções para escolher apenas uma delas: “(i) sua maluta; (ii) seu malute; (iii) sua malute; (iv) seu maluto”. Por outro lado, para as três questões que envolviam pseudoverbos na forma derivada, os informantes, por exemplo, teriam, para a frase “Nos aposentos da casa, devido ao excesso de umidade, somente com ZENETO mantemos a casa limpa. Se esquecemos de , sofreremos com alergias o ano inteiro“, três opções para escolher apenas uma delas: (i) zenetar; (ii) zeneter; (iii) zenetir.

Os participantes deste experimento foram convidados por e-mail ou postagem em redes sociais. O critério para participação no experimento foi estar matriculado em algum curso de graduação dentro do território brasileiro. Antes de sua participação no experimento, cada indivíduo teve acesso ao Termo de Consentimento Livre e Esclarecido (Apêndice B), informando que conheciam os objetivos da pesquisa, que estavam cientes de que leriam frases com palavras inventadas, que nenhum dado pessoal seria revelado em nenhuma instância da pesquisa e que, por serem voluntários, poderiam desistir de participar da pesquisa em qualquer etapa e, tendo com essa decisão, seus dados excluídos da análise. O teste não teve custos ou

³⁷ A escolha de alunos de graduação se deve ao fato de limitarmos nossa amostra de informantes àqueles com formação escolar mínima de Ensino Médio completo e também ao fato de contarmos em nossas redes de contato com indivíduos que cursam o Ensino Superior.

benefícios financeiros para os participantes. Além disso, a participação na pesquisa não representou nenhum risco moral aos participantes.

4.7 COMPARAÇÃO ENTRE RESULTADOS: LÉXICO DICIONARIZADO, LÉXICO INSTITUCIONALIZADO E LÉXICO POTENCIAL

Na etapa 7, comparamos as formas realizadas pelos informantes no experimento com pseudopalavras (léxico potencial) e os padrões mais frequentes observados no dicionário (léxico dicionarizado) e no *Corpus Brasileiro* (léxico institucionalizado).

5 RESULTADOS

Neste capítulo apresentamos, discutimos e comparamos os resultados levantados a partir da consulta ao léxico dicionarizado, representado pelos dicionários eletrônicos *Aurélio* e *Houaiss*, e ao léxico institucionalizado, no *Corpus Brasileiro*, e do léxico potencial, com a aplicação do experimento online com pseudopalavras.

5.1 LÉXICO DICIONARIZADO: DICIONÁRIOS ELETRÔNICOS *AURÉLIO* E *HOUAISS*

Em primeiro lugar, procedeu-se à coleta de substantivos derivados de verbos por meio do processo de derivação regressiva. Para isso, utilizamos o Dicionário Eletrônico *Aurélio*, versão 5.0.40. A escolha desse dicionário é justificada pelo fato de oferecer o recurso de aplicação de “filtros” (máscaras), o que possibilitou reduzir a lista de entradas a substantivos deverbais terminados em “a”, “e” e “o”.

A partir desse filtro, foram analisadas individualmente essas palavras para selecionar apenas os casos de substantivos deverbais via derivação regressiva, ou seja, apenas os casos definidos pelo dicionário como nomes deverbais. Ao selecionar as opções de filtro “Procurar no texto do verbete” e “Substantivo”, foram encontrados 3.377 nomes deverbais.

Conforme Schwindt (2018), o Dicionário Eletrônico *Aurélio* conta com 17.049 *types* de substantivos (excluindo-se compostos, derivados de outros nomes e elementos de locuções). Desse total, 39,3% são de palavras terminadas em *a*, 23% em *o* e 14,8% em *e* (os outros casos se referem a biforme *a/o* átono, por exemplo, *menino/a*, com 1,5 %, outros átonos, como *móvel*, com 2,9%, e tônicos como *jabá*, *anel*, com 18,7%), como podemos ver na Tabela 1.

Tabela 1 - Distribuição da vogal final dos substantivos no Dicionário Eletrônico *Aurélio*

Vogal final do substantivo	% de 17.049 types - Dicionário Eletrônico <i>Aurélio</i>
a	39,3
o	23
e	14
outros casos	23,7

Fonte: Adaptado de Schwindt, 2018, p. 753.

Do total de 3.377 verbetes derivados de verbos, selecionamos apenas os casos indicados pelo dicionário como derivados de verbos. Assim, foram obtidos 979 nomes deverbais, conforme a distribuição entre as três vogais finais.

Tabela 2 - Distribuição das três vogais finais nos nomes deverbais

Vogal final	%	Ocorrências
o (p. ex., abandon <u>o</u>)	60	581
a (p. ex., perd <u>a</u>)	26	255
e (p. ex., ataqu <u>e</u>)	14	143
TOTAL		979

Fonte: Elaboração própria.

Quando comparamos as Tabelas 1 e 2, percebemos que a distribuição geral de substantivos terminados em *o* e *a* é distinta da que observamos para os substantivos deverbais por derivação regressiva: há uma frequência relativa muito superior dos nomes deverbais terminados em *o* (60%) em comparação à categoria geral dos substantivos (23%), em que os substantivos terminados em *a* são mais frequentes (39,3%). Mais adiante será possível perceber que a frequência elevada de nomes deverbais terminados em *o* se mantém nesse mesmo percentual elevado mesmo depois de serem excluídos verbetes por diversos critérios, como regionalismo, *homofonia acidental*, alternância da vogal final e registro em apenas um dos dois dicionários consultados.

Abaixo temos uma Tabela (3) que apresenta a distribuição das três conjugações nos verbetes do dicionário eletrônico Aurélio. Conforme o esperado dada a grande produtividade dos verbos da primeira conjugação, temos 89% de dados pertencentes a esse grupo.

Tabela 3 - Distribuição das conjugações verbais nos nomes deverbais em geral

Conjugação verbal	%	Ocorrências
ar (p. ex., abandon <u>o</u> >abandon <u>ar</u>)	89	11.314
er (p. ex., perd <u>a</u> >perder <u>er</u>)	6	716
ir (p. ex., fug <u>a</u> >fugir <u>ir</u>)	5	679
TOTAL		12.709

Fonte: Elaboração própria.

Em seguida, esses 979 vocábulos deverbais foram consultados no Dicionário Eletrônico *Houaiss*, versão 3, a fim de corroborar a existência dessas formas e seu caráter verbal. Ao todo, 26 vocábulos foram eliminados da lista de entradas coletada por meio do Dicionário *Aurélio*: 21 palavras terminadas em *o*; e 5 palavras terminadas em *e*. Nenhuma palavra terminada em *a* foi excluída.

Conforme vemos a seguir na Tabela 4, a segunda análise da lista de vocábulos, agora um total de 953, retirou todos os casos (identificados pelos dicionários, em um deles ou em

ambos) de regionalismo, formas com *homofonia acidental* e palavras que registravam alternância no uso da vogal final (por exemplo, “desmama” e “desmame”). Nessa etapa, foram excluídos 276 vocábulos.

Tabela 4 - Palavras encontradas nos dois dicionários

Vogal final	%	Ocorrências
o (p. ex., abandon <u>o</u>)	59	560
a (p. ex., perd <u>a</u>)	27	255
e (p. ex., ataqu <u>e</u>)	14	138
TOTAL		953

Fonte: Elaboração própria.

Na Tabela 5 abaixo, temos a frequência de nomes deverbais após a exclusão de 57 palavras: 30 palavras terminadas em *o*, 21 palavras terminadas em *a* e 6 palavras terminadas em *e*.

Tabela 5 - Palavras sem regionalismos

Vogal final	%	Ocorrências
o (p. ex., abandon <u>o</u>)	59	530
a (p. ex., perd <u>a</u>)	26	234
e (p. ex., ataqu <u>e</u>)	15	132
TOTAL		896

Fonte: Elaboração própria.

A seguir, na Tabela 6, encontramos a frequência de nomes deverbais após serem excluídas todas as formas com *homofonia acidental*, ou seja, apresentarem uma mesma grafia para palavras com sentido totalmente diverso. Foram excluídas 20 palavras terminadas em *o*, 26 palavras terminadas em *a* e 8 palavras terminadas em *e*.

Tabela 6 - Palavras sem formas com homofonia acidental

Vogal final	%	Ocorrências
o	60	510
a	25	208
e	15	124
TOTAL		842

Fonte: Elaboração própria.

Na Tabela 7 abaixo, vemos a frequência resultante da exclusão de nomes deverbais que permitem mais de uma vogal final como *desmama* e *desmame*. Foram excluídas 43 palavras terminadas em *o*, 79 palavras terminadas em *a* e 43 palavras terminadas em *e*.

Tabela 7 - Palavras sem formas alternantes

Vogal final	%	Ocorrências
o	69	467
a	19	129
e	12	81
TOTAL		677

Fonte: Elaboração própria.

Desse modo, ao final da etapa de consulta aos dicionários, restavam 677 substantivos deverbiais selecionados segundo os seguintes critérios:

- (i) listados nos dois dicionários eletrônicos pesquisados, Aurélio e Houaiss;**
- (ii) definidos como nomes derivados de verbos;**
- (iii) de linguagem vernacular, em oposição a casos de regionalismo, forma com homofonia acidental e alternância na forma de registro da vogal final (a, o e e).**

Esse resultado diverge do apontado por Schwindt (2018) para os substantivos no léxico dicionarizado do português brasileiro, em que houve uma frequência de 39,3% palavras terminadas em *a* e 23% para palavras terminadas em *o*. Assim, temos duas tendências opostas, uma mais geral referente aos substantivos no português brasileiro, palavras terminadas em *a*, e outra tendência mais restrita referente aos nomes deverbiais via derivação regressiva, palavras terminadas em *o*.

5.1.1 Datação das palavras

O Dicionário *Houaiss* apresenta para uma parte de seus vocábulos a informação sobre a data em que foi registrado pela primeira vez em um documento escrito. Com base nessa informação, estabelecemos uma subamostra em que foi organizada a distribuição das vogais finais por uma linha do tempo. Do total de 953 formas corroboradas pelo *Houaiss*, havia informação sobre a datação do primeiro registro escrito em 657 palavras (69%). Pode-se observar essa distribuição temporal na Tabela 8 abaixo.

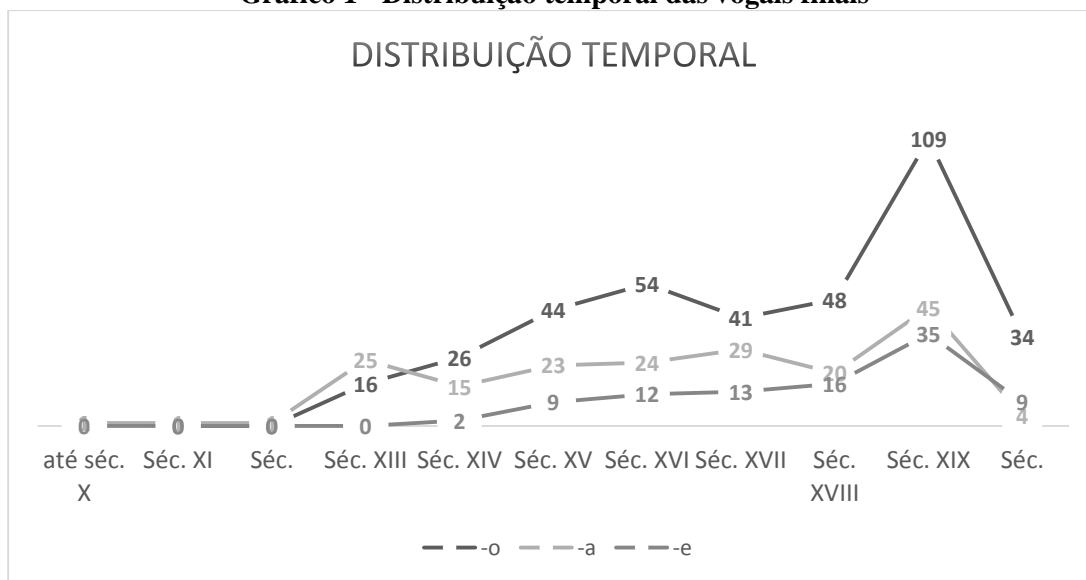
Tabela 8 - Distribuição das vogais finais pela linha do tempo

	até séc. X	séc. Xi	séc. xii	séc. xiii	séc. xiv	séc. Xv	séc. Xvi	séc. xvii	séc. xviii	séc. xix	séc. xx	Total
o	01	00	00	16	26	44	54	41	48	109	34	373
a	01	01	01	25	16	23	24	29	20	45	04	189
e	00	00	00	00	02	09	12	13	16	35	09	96
Total	02	01	01	41	44	76	90	83	84	189	47	657

Fonte: Elaboração própria.

O Gráfico 1 a seguir torna mais explícito ainda o padrão de distribuição quase absolutamente uniforme: em primeiro lugar, palavras terminadas em *o*, em segundo lugar, palavras terminadas em *a*, e em terceiro lugar, palavras terminadas em *e*.

Gráfico 1 - Distribuição temporal das vogais finais



Fonte: Elaboração própria.

A tendência de predomínio dos nomes deverbais terminados em *o* para o fenômeno da derivação regressiva parece não ser um evento novo na língua portuguesa, conforme aponta nossa breve análise de palavras com a informação de datação do primeiro registro escrito no Dicionário Eletrônico *Houaiss*.

5.1.2 Análise de Regionalismos, Formas com Homofonia Acidental, Formas Alternantes e Formas Não Listadas

Nesta seção tratamos dos casos de vocábulos excluídos devido à sua natureza: regionalismos, formas com *homofonia acidental* e formas alternantes.

Nesta etapa de pesquisa nos dicionários, percebeu-se uma distribuição constante da frequência de ocorrência das vogais finais nesta ordem: *o*, *a* e *e*. Esse padrão pode ser observado na Tabela 9 a seguir.

Observa-se, na Tabela 9, que apenas quando se excluem regionalismos, formas com homofonia acidental e formas alternantes há uma oscilação na distribuição das três vogais finais. Por isso, procedeu-se a uma análise mais detalhada dessas formas excluídas para levantar possíveis informações importantes para entender o fenômeno da derivação regressiva que

poderiam estar sendo ignoradas nesse processo de exclusão. Apenas foram mantidas fora dessa análise as formas com homofonia acidental porque os recursos de filtro no *Corpus Brasileiro* não distinguem substantivos deverbiais de outros tipos de substantivos. Assim, para uma forma como *guarda*, não poderíamos ter certeza se estávamos diante de apenas casos de derivação regressiva do verbo *guardar* ou de um substantivo primitivo que indica a profissão de *guarda* (sinônimo de *vigilante*).

Tabela 9 - Comparação entre critérios de organização dos dados

	Auréli o %	Ocorrê ncias	Houais s %	Ocorrê ncias	Exclusõ es %	Ocorrê ncias	Dataçã o %	Ocorrê ncias
o	60	581	59	560	69	467	57	373
a	26	256	27	256	19	129	29	189
e	14	143	14	138	12	81	14	96
TOTAL		979		954		677		658

Fonte: Elaboração própria.

Em primeiro lugar, debruçamo-nos sobre os casos de regionalismo.

Como apresentado na Tabela 10 abaixo, a distribuição das vogais finais replica, com variação, o padrão observado no léxico.

Tabela 10 - Distribuição de vogais finais nos casos de regionalismo

Vogal final	%	Ocorrências
o	53	30
a	37	21
e	10	6
TOTAL		57

Fonte: Elaboração própria.

A partir do levantamento da frequência de *tokens* no *Corpus Brasileiro*, para palavras terminadas em *o*, 7 palavras apresentaram no mínimo uma ocorrência no *Corpus Brasileiro*, como vemos na Tabela 11.

Tabela 11 - Regionalismos: frequência para vogal final o no Corpus Brasileiro

(continua)

Regionalismo	%	Ocorrências
rejeito	86	823
calço	10	92
remelexo	3	31
pipoco	0,4	5

Tabela 11 - Regionalismos: frequência para a vogal final no Corpus Brasileiro

(conclusão)

Regionalismo	%	Ocorrências
costeio	0,3	3
saçarico	0,2	2
aforro	0,1	1
TOTAL		957

Fonte: Elaboração própria.

Na Tabela 11 acima, vimos que o regionalismo mais frequente é “rejeito”. Esse dado não pode ser incluído na análise geral porque pode se referir tanto à forma derivada do verbo rejeitar que mantém a semântica do verbo como “ato de rejeitar” quanto a um objeto, que pode ser substituído pelo vocábulo “lixo”. As 6 restantes podem ser incluídas na análise geral dos dados.

Os outros 24 regionalismos restantes não foram encontrados no *Corpus Brasileiro*: *abodego, açulo, adolo, aperreio, barrufo, cabeio, campeio, costeio, empreito, empurro, escaramuceio, escoo, estrouxo, garreio, gateio, palanqueio, paleio, pipoco, rasqueteio, repasso, rusto, vacilo, verdeio, verdejo e zanguizarreio*.

Conforme vemos na Tabela 12, para as palavras terminadas em *a*, 6 palavras apresentaram no mínimo uma ocorrência no *Corpus Brasileiro*.

Tabela 12 - Regionalismos: frequência para vogal final a no Corpus Brasileiro

Regionalismo	%	Ocorrência
engorda	61	1.370
cata	22	502
despesca	9	201
cabruca	5	115
tosa	2	43
monda	1	10
TOTAL		2.241

Fonte: Elaboração própria.

Os outros 15 regionalismos terminados em *a* não apresentaram nenhuma ocorrência no *corpus*: *beija, cabroca, cácea, colha, derriça, derruba, empreita, encerra, encruza, enrola, manja, paquera, peja, seva e subestima*.

Na Tabela 13, vemos que, para as palavras terminadas em *e*, foram encontrados quatro regionalismos no *Corpus Brasileiro*. Apenas as formas *desaguache* e *despenque* não foram encontradas no *Corpus Brasileiro*.

Tabela 13 - Regionalismos: frequência para vogal final e no Corpus Brasileiro

Regionalismo	%	Ocorrência
aparte	98	9444
desbutonde	1	142
engorde	0,4	40
arribe	0,6	1
TOTAL		9.627

Fonte: Elaboração própria.

Em relação às formas que apresentam alternância na emergência da vogal final na forma derivada, conforme apresentado abaixo na Tabela 14, para as palavras terminadas em *o*, foram encontradas 19 palavras com frequência no *Corpus Brasileiro*.

Tabela 14 - Formas com alternância: frequência para vogal final o no Corpus Brasileiro

Forma com alternância	%	Ocorrência
avanço	88	31.995
troco	6	1.868
arrasto	2	754
ressalto	1,6	587
lanço	1,2	454
reclamo	0,6	260
talho	0,4	154
arranco	0,2	74
desmancho	0	20
esparramo	0	16
desfruto	0	15
ameaço	0	12
recolho	0	12
saracoteio	0	5
soçobro	0	5
deslastro	0	4
desgarro	0	3
traspasso	0	3
encalho	0	2
TOTAL		36.243

Fonte: Elaboração própria.

As outras 24 formas não foram encontradas no *corpus*: *alcanço, apanho, decalco, derrengo, descarrego, desdobro, desencalho, desengasgo, desforro, desgasto, deslustro, discordo, empacho, empeno, encaixo, engasgo, entrelaço, enxugo, escaldo, farfalho, pago, rebusco, transporte e trasfego*.

Para as palavras terminadas em *a*, como vemos abaixo na Tabela 15, encontramos 21 delas no *Corpus Brasileiro*.

Tabela 15 - Formas com alternância: frequência para vogal final a no Corpus Brasileiro

Forma com alternância	%	Ocorrência
conta	79	128.026
quebra	11	18.805
trilha	6	9.780
desmama	2	2.688
saca	0,7	1.415
avança	0,3	532
recolha	0,2	381
talha	0,2	287
desforra	0,1	205
conjura	0,1	172
debulha	0,1	158
empreita	0	96
apanha	0	95
rebaixa	0	81
tosa	0	43
gira	0	21
achega	0	16
escalda	0	9
pernoita	0	8
respiga	0	4
rebusca	0	2
TOTAL		162.824

Fonte: Elaboração própria.

As 15 restantes não apresentaram nenhuma ocorrência no *Corpus Brasileiro*: *acossa*, *aguarda*, *arranca*, *desfruta*, *desmancha*, *destêmpera*, *encarna*, *ensaca*, *experimenta*, *grita*, *pousa*, *recua*, *soçobra*, *sossega* e *trasfega*.

Como vemos a seguir na Tabela 16, para as palavras terminadas em *e*, encontramos 21 no Corpus Brasileiro.

Tabela 16 - Formas com alternância: frequência para vogal final e no Corpus Brasileiro

(continua)

Forma com alternância	%	Ocorrência
transporte	55	60.480
alcance	22	24.488
desgaste	10	10.526
lance	5	5.820
derrame	3	3.038
levante	2	1.796
encaixe	1	1.193
desfrute	0,3	406
arranque	0,3	329
pernoite	0,2	242
encalhe	0,2	241

Tabela 16 - Formas com alternância: frequência para a vogal final e no Corpus Brasileiro
(conclusão)

Forma com alternância	%	Ocorrência
tique	0	187
descasque	0	95
traspasse	0	68
retarde	0	59
derrube	0	56
soque	0	53
ressalte	0	20
desgarre	0	8
desbarate	0	4
esparrame	0	2
TOTAL		109.111

Fonte: Elaboração própria.

As outras 22 palavras não foram encontradas no *corpus*: *aceite, arraste, avance, derranque, derrengue, desdobre, desencalhe, desengasgue, deslastre, deslustre, desmame, emboque, encarne, endosse, engasgue, ensaque, entrelace, esgote, pernoute, rebaixe, rebusque e repasse*.

Na Tabela 17, temos as formas alternantes com frequência no Corpus Brasileiro. Das nove formas encontradas no *corpus*, 7 apresentam o seguinte padrão: alternância entre as vogais *o* e *e*, com frequência maior da vogal *e*. Apenas a forma *encalho/encalhe* apresentou uma frequência maior para a vogal final *o*.

Tabela 17 - Distribuição das formas alternantes nas três vogais finais: types

	o	a	e
arranc-	74	0	329
desfrut-	15	0	406
desgarr-	3	0	8
encalh-	2	0	241
esparram-	16	0	2
lanc-	454	0	5820
ressalt-	587	0	20
talh-	154	287	0
traspas-	3	0	68

Fonte: Elaboração própria.

A única forma que apresentou uma alternância *o/a* foi *talho/talha*, com maior frequência de vogal final *a*. Esse resultado parece apontar para o fato de que quando há alternância, a frequência maior é de palavras com vogal final *e*, o que contraria o padrão dos nomes deverbiais por derivação regressiva, que distribuem a frequência de ocorrência na ordem *o, a e e*.

Na Tabela 18, a seguir, temos o registro de datação da primeira ocorrência documentada de algumas das 9 formas alternantes. Há apenas duas formas que têm datação para as três vogais finais: *arranc-* e *talh-*. Não é possível encontrar um padrão para essas duas formas: quando a vogal *e* é mais frequente, caso de *arranc-*, *arranca* é a forma mais nova e *arranco*, a forma com o primeiro registro mais antigo; quando a vogal mais frequente é *a*, *talha* é a forma mais velha e *talhe*, a mais nova.

Tabela 18 - Distribuição das formas alternantes nas três vogais finais: tokens

	o	a	e
arranc-	1611	1899	1802
desfrut-	/	/	1881
desgarr-	/	/	/
encalh-	/	/	1813
esparram-	1899	/	/
lanc-	1364	/	1676
ressalt-	1635	/	1899
talh-	1365	1265	1501
traspas-	1390	/	1817

Fonte: Elaboração própria.

Além disso, há três formas com datação para *o* e *e*: *lanc-*, *ressalt-* e *traspas-*. Em todas elas, *e* é a forma mais nova. Isso pode ser um fator relevante para a alta frequência de palavras terminadas em *e* quando comparadas com sua contraparte em *o*: quanto mais recente for o registro da palavra, quando comparada com sua alternante em *o*, maior a probabilidade de ter uma frequência relativa maior. É claro que seria necessária uma pesquisa em linguística histórica mais ampla para determinar se essa é uma regra válida para os nomes deverbais com alternância em *o/e*.

Das formas não encontradas no dicionário Houaiss, para as palavras terminadas em *o*, 12 palavras foram encontradas no Corpus Brasileiro, como é apontado na Tabela 21.

Tabela 19 - Formas não listadas no Houaiss: frequência da vogal final o no Corpus Brasileiro

(continua)

Forma não listada	%	Ocorrência
trato	48	14.585
relevo	31	9.485
coco	16	4.982
retiro	3	854
tropeço	2	496
aboio	0	63
apresto	0	12
restilo	0	4

Tabela 19 - Formas não listadas no Houaiss: frequência da vogal final o no Corpus Brasileiro (conclusão)

Forma não listada	%	Ocorrência
arregaço	0	1
arrenego	0	1
derriço	0	1
pareio	0	1
TOTAL		30.485

Fonte: Elaboração própria.

As palavras *relevo*, *coco* e *retiro*, por poderem se referir tanto a ações como a objetos ou lugares, foram excluídas da análise geral.

As nove restantes não foram encontradas no *Corpus Brasileiro*: *adergo*, *arrebato*, *escancarar*, *reverbero*, *rufio*, *tamborilo*, *tartareio*, *tenetio* e *tonteio*.

Para as formas não listadas terminadas em *e*, como vemos na Tabela 20, das cinco formas encontramos duas no *Corpus Brasileiro*, *pique* e *requeime*. As três formas restantes não encontradas foram: *bebe*, *calque* e *tilinte*.

Tabela 20 - Formas não listadas no Houaiss: frequência da vogal final e no Corpus Brasileiro

Forma não listada	%	Ocorrência
pique	99,9	982
requeime	0,1	1
TOTAL		983

Fonte: Elaboração própria.

Assim, conforme registrado na Tabela 21, foram novamente incluídas devido à ocorrência comprovada no *Corpus Brasileiro* 82 palavras na análise geral, o que resulta em um total de 759 dados.

Tabela 21 - Distribuição das palavras incluídas para as três vogais finais

	Número de palavras incluídas (o)	Número de palavras incluídas (a)	Número de palavras incluídas (e)	TOTAL
Regionalismos	6	6	4	16
Formas com alternância	18	18	19	55
Formas não listadas	9	0	2	11
TOTAL	33	24	25	82

Fonte: Elaboração própria.

Na Tabela 22, a seguir, temos a distribuição de nomes deverbiais selecionados a partir de verbetes encontrados nos dicionários *Aurélio* e *Houaiss* e, desses nomes deverbiais selecionados, os nomes encontrados no *Corpus Brasileiro*.

Tabela 22 - Distribuição das vogais finais no dicionário e no Corpus Brasileiro

Vogal final	dicionário %	Ocorrências	Corpus Brasileiro %	Ocorrências
o	66	500	59	292
a	20	153	26	128
e	14	106	15	74
TOTAL		759		494

Fonte: Elaboração própria.

Podemos perceber que a distância entre as palavras terminadas em *o* e *a* diminui no *corpus*, porém ainda é uma diferença bastante significativa quando comparada com a distribuição geral dos substantivos, com 42% para palavras terminadas em *o* e 40% para palavras terminadas em *a*.

5.1.3 Resultado da codificação segundo os grupos de fatores linguísticos

Nesta seção, apresentamos os resultados da codificação dos dados conforme as variáveis linguísticas discutidas no Capítulo 5, sobre nossos procedimentos metodológicos.

5.1.3.1 Vogal temática verbal

Na Tabela 23, temos a distribuição das frequências no léxico dicionarizado para as três conjugações verbais. Como podemos ver a grande maioria dos verbetes pertence a nomes derivados de verbos da primeira conjugação. Todas as vogais finais apresentam formas derivadas de verbos.

Tabela 23 - Fator Vogal Temática Verbal: distribuição das vogais finais

	o %	Ocorrências	a %	Ocorrências	e %	Ocorrências
primeira conjugação	98	286	94	120	95	70
segunda conjugação	1,4	4/	5,4	7	5	4
terceira conjugação	0,6	2	0,6	1	0	0
TOTAL		292		128		74

Fonte: Elaboração própria.

Os nomes derivados de segunda conjugação são: (*o*) *socorro, sorvo, remelexo e recolho*; (*a*) *contenda, encolha, escolha, foda, penhora, revenda e recolha*; (*e*) *abate, combate, debate e rebate*. Somente para palavras terminadas em *o* e *a* encontramos nomes derivados de verbos da terceira conjugação: respectivamente, *consumo, tino, e fuga*.

5.1.3.2 Vogal temática

Na Tabela 24, temos a variável vogal temática, que replica os resultados da variável conjugação verbal: palavras terminadas em *o* e *e* mostram uma tendência à alteração da vogal temática na passagem de verbo a nome por derivarem de verbos da primeira conjugação, e as palavras terminadas em *a*, em sua extensa maioria, preservam a vogal temática da primeira conjugação.

Tabela 24 - Fator Vogal Temática: distribuição das vogais finais

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
alteração da VT	100	292	6	8	94	70
manutenç ão da VT	0	0	94	120	6	4
TOTAL		292		128		74

Fonte: Elaboração própria.

Esse resultado reflete a alta frequência de nomes derivados de verbos da primeira conjugação (476) associada à alta frequência de nomes deverbais terminados em *o* (292), que não tem nenhum correspondente nas vogais temáticas com a mesma qualidade vocálica, em oposição aos casos das vogais finais *a* e *e*, que correspondem à primeira e à segunda conjugações verbais.

5.1.3.3 Complexidade morfológica

A Tabela 25 apresenta a distribuição de ocorrências no léxico dicionarizado da variável complexidade morfológica. Para as palavras terminadas em *o* e *a*, a frequência maior é de palavras sem afixos. Já o contrário ocorre para palavras terminadas em *e*: 61% dos casos são de palavras com afixos.

Tabela 25 - Fator Complexidade Morfológica: distribuição das vogais finais

	o	Ocorrências	a	Ocorrências	e	Ocorrências
	%		%		%	
simples	63	183	78	100	39	29
complexa	37	109	22	28	61	45
TOTAL		292		128		74

Fonte: Elaboração própria.

Dessas palavras complexas terminadas em *e*, temos a seguinte distribuição de prefixos, em ordem decrescente, conforme os dados da Tabela 26.

Tabela 26 - Distribuição dos prefixos para a vogal final e

Prefixo	%	Ocorrência
des-	33	15
re-	29	13
en-	15	7
a-	4,4	2
de-	4,4	2
in-	2,2	1
per-	2,2	1
trans-	2,2	1
tras-	2,2	1
tre-	2,2	1
tres-	2,2	1
TOTAL		45

Fonte: Elaboração própria.

Acima vemos que os prefixos mais frequentes no léxico para palavras terminadas em *e* são *des-* (33%) e *re-* (29%). É interessante comparar esses resultados com os prefixos encontrados nas palavras terminadas em *o* e *a*.

Conforme a Tabela 27, os prefixos *des-* e *re-* também são os mais frequentes para as palavras terminadas em *o*. Além disso, diferentemente dos nomes terminados em *e*, aqui temos também a presença dos sufixos *-ejo*, *-eio* e *-isco*.

Tabela 27 - Distribuição dos prefixos para a vogal final o

(continua)

Prefixo	%	Ocorrência
des-	38	42
re-	16	18
a-	12	13
trans-	4	4
en-	3	3
es-	3	3
con-	2	2

Tabela 27 - Distribuição dos prefixos para a vogal final o

(continuação)

Prefixos	%	Ocorrência
sobre-	1	1
Sufixos	%	Ocorrência
-ejo	10	11
-eio	9	10
-isco	2	2
TOTAL		109

Fonte: Elaboração própria.

Para os nomes terminados em *a*, na Tabela 28, também temos os prefixos *re-* e *des-* como os mais frequentes (com uma distribuição inversa em comparação com os outros dois casos acima). Proporcionalmente, os nomes terminados em *a* são os que apresentam o menor percentual de palavras morfológicamente complexas, ou seja, com a presença de afixos.

Tabela 28 - Distribuição dos prefixos para a vogal final a

Prefixo	%	Ocorrência
re-	40	11
des-	33	9
a-	15	4
en-	4	1
es-	4	1
per-	4	1
TOTAL		27

Fonte: Elaboração própria.

Resumindo, essa variável parece indicar uma tendência de nomes deverbais terminados em *e* apresentarem os prefixos *des-* e *re-* (que são prefixos com frequência alta em geral). Ademais, os nomes deverbais terminados em *o* e *a* apresentam uma tendência a ter uma estrutura morfológica simples, com a diferença que *a* apresenta uma tendência maior a apresentar esse padrão e que os nomes terminados em *o* são o único tipo que apresenta em alguns casos a presença de sufixos (*-ejo*, *-eio* e *-isco*).

5.1.3.4 Vogal média tônica

Quanto à variável vogal média tônica, a Tabela 29 indica que a maioria dos nomes deverbais não apresentou esse tipo de vogal, com destaque para a mais baixa frequência no caso de nomes terminados em *e*, com apenas 11% (8 casos) de seu total (74).

Tabela 29 - Fator Vogal Média Tônica: distribuição das vogais finais

	o %	Ocorrências	a %	Ocorrências	e %	Ocorrências
presença	39	113	41	52	11	8
ausência	61	179	59	76	89	66
TOTAL		292		128		74

Fonte: Elaboração própria.

Os nomes terminados em *e* com vogal média tônica são: *corte*, *enfoque*, *entrechoque*, *informe*, *remonte*, *retoque*, *engorde* e *soque*.

5.1.3.5 Gênero da palavra

Quanto à variável gênero da palavra, conforme dados da Tabela 30, embora o português brasileiro tenha mais nomes femininos do que masculinos, há uma alta frequência de nomes deverbais masculinos.³⁸

Tabela 30 - Fator Gênero da Palavra: distribuição das vogais finais

	o %	Ocorrências	a %	Ocorrências	e %	Ocorrências
masculino	100	292	0	0	100	74
feminino	0	0	100	128	0	0
TOTAL		292		128		74

Fonte: Elaboração própria.

Isso se deve principalmente ao fato de haver um predomínio de nomes deverbais terminados em *o*, que, nesse caso, está relacionado ao gênero gramatical masculino.

5.1.3.6 Número de sílabas

Para a variável número de sílabas, conforme vemos na Tabela 31, a maioria dos nomes deverbais apresenta uma estrutura trissilábica. O que chama atenção na distribuição dessa variável é a alta frequência relativa de dissílabos terminados em *a* (33%), o que pode indicar uma tendência relacionada a essa variável.

³⁸ Esse resultado contraria o observado por Schwindt (2018), que demonstrou haver mais formas femininas do que masculinas em formas nominais do PB.

Tabela 31 - Fator Número de Sílabas: distribuição das vogais finais

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
monossíla bo	0	0	0	0	0	0
dissílabo	12	36	33	43	9	7
trissílabo	66	191	56	71	76	56
polissílabo	22	65	11	9	15	11
TOTAL		292		128		74

Fonte: Elaboração própria.

Segundo Viaro e Guimarães-Filho (2007), com base em 150.875 *types* extraídos do Dicionário Houaiss e submetidos a programas computacionais da plataforma *MatLab*, há no léxico dicionarizado do português brasileiro a seguinte distribuição de palavras segundo a quantidade de sílabas:

- (i) **monossílabos**: 546
- (ii) **dissílabos**: 11.712
- (iii) **trissílabos**: 36.790
- (iv) **polissílabos**: 101.828

Acima, podemos ver que há no léxico mais polissílabos, seguidos de trissílabos, o que não se reflete na distribuição vista na Tabela 31, em que há uma frequência maior de trissílabos. Porém, a distribuição geral se refere ao léxico como um todo, não somente à classe dos substantivos. Assim, não é possível afirmar com convicção se esse resultado do fator número de sílabas reflete um padrão específico dos nomes deverbais ou um padrão mais geral da classe dos substantivos.

5.1.3.7 Onset da última sílaba

Em relação à variável *onset* da última sílaba, como aponta a Tabela 32, a maioria quase absoluta dos nomes deverbais apresenta *onset* nessa posição silábica.

Tabela 32 - Fator Onset da Última Sílaba: distribuição das vogais finais

	o %	Ocorrências	a %	Ocorrências	e %	Ocorrências
presença	86	252	97	124	100	74
ausência	14	40	3	4	0	0
TOTAL		292		128		74

Fonte: Elaboração própria.

Podemos destacar o leve desvio desse padrão por parte dos nomes terminados em *o*, que apresentou 40 casos (14%) de ausência do *onset* na sílaba final. Isso se deve ao fato de haver em todos esses casos a sequência final -io (em muitos casos, resultante do acréscimo da vogal final *o* e a consequente emergência da epêntese -i- precedendo essa vogal, como por exemplo no nome verbal *bloqueio*).

5.1.3.8 Contexto fonológico precedente (modo de articulação)

No que se refere ao modo de articulação do contexto fonológico precedente, na Tabela 33 podemos observar que as plosivas são o contexto mais frequente para a derivação regressiva.

Tabela 33 - Fator Contexto Fonológico Precedente (modo de articulação): distribuição das vogais finais

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
nasal	9	27	9	11	5	4
plosiva	29	86	44	57	76	56
fricativa	26	75	17	22	16	12
lateral	9	27	9	12	3	2
tepe	9	27	15	19	0	0
vogal	18	50	6	7	0	0
TOTAL		292		128		74

Fonte: Elaboração própria.

Aqui podemos destacar o aumento inversamente proporcional da frequência de plosivas conforme nos movemos dos nomes verbais mais frequentes (terminados em *o*), 29% dos dados em *o*, para os nomes verbais menos frequentes (terminados em *e*), 76% dos dados em *e*, com os nomes terminados em *a* ocupando uma posição intermediária, 44% dos dados em *a*.

5.1.3.9 Contexto fonológico precedente (ponto de articulação)

Em relação à variável ponto de articulação do contexto fonológico precedente, os dados da Tabela 34 indicam que há num menor grau a relação inversa entre frequência da vogal final dos nomes verbais e distribuição do fator mais frequente, contexto alveolar.

Tabela 34 - Fator Contexto Fonológico Precedente (ponto de articulação): distribuição das vogais finais

	o %	Ocorrências	a %	Ocorrências	e %	Ocorrências
bilabial	4	13	6	7	9	7
labiodental	3	2	2	3	0	0
alveolar	44	128	52	67	55	41
palatal	18	52	13	17	4	3
velar	14	40	21	27	32	23
vogal alta	16	49	5	6	0	0
vogal média	0,3	1	0,8	1	0	0
vogal baixa	0	0	0	0	0	0
TOTAL		292		128		74

Fonte: Elaboração própria.

Aqui cabe destacar também o alto índice de contexto velar para os nomes terminados em *e* (32%) e a inexistência de vogais nesse contexto.

5.2 LÉXICO INSTITUCIONALIZADO: *CORPUS BRASILEIRO*

Por meio do *Corpus Brasileiro*, que conta com 938.7 milhões de palavras, foi consultada a frequência lexical dos 759 vocábulos selecionados nesta pesquisa a partir da exclusão de alguns de casos de regionalismo, formas alternantes e formas ambíguas. Desse total, 265 vocábulos apresentaram frequência zero. Restaram, assim, 494 vocábulos com registro no *corpus* (lista de palavras com a respectiva frequência lexical de cada vocábulo em Apêndice G).

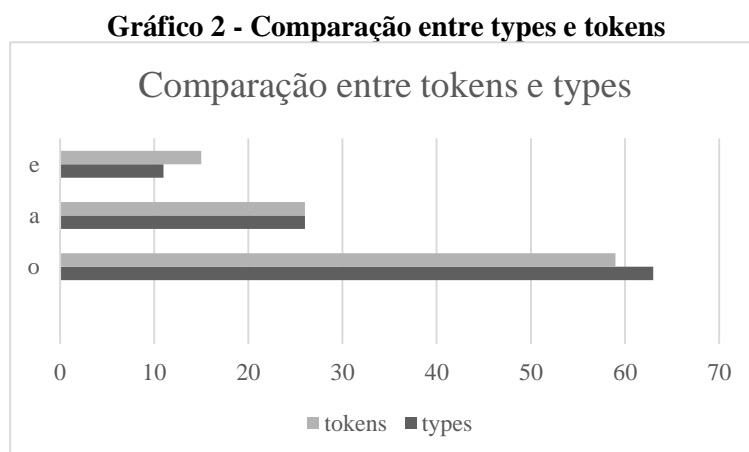
Na Tabela 35 podemos perceber uma pequena variação entre a distribuição de *tokens* e *types* no *Corpus Brasileiro* o que pode refletir na comparação entre as frequências dos grupos de fatores linguísticos para *tokens* e *types*, como veremos mais a frente.

Tabela 35 - Comparação entre tokens e types: distribuição das vogais finais

Vogal final	Tokens %	Ocorrências	Types %	Ocorrências
o	63	2.794.606	59	292
a	26	1.152.744	26	128
e	11	463.053	15	74
TOTAL		4.410.403		494

Fonte: Elaboração própria.

Abaixo, no Gráfico 2 podemos perceber de forma mais nítida a uniformidade na distribuição das três vogais para types e tokens. Esse resultado corrobora a perspectiva da Teoria de Exemplos, ao demonstrar como um padrão de uso (*tokens*) é replicado no padrão gramatical da língua (*types*).



Fonte: Elaboração própria.

A seguir apresentamos o resultado da distribuição de ocorrências para as variáveis linguísticas em análise no léxico institucionalizado do português brasileiro, da mesma forma que apresentamos a distribuição para o léxico dicionarizado do PB.

5.2.1 Vogal temática verbal

Em relação à variável vogal temática verbal, a Tabela 36 apresenta a frequência no léxico institucionalizado para as três conjugações verbais. Como podemos ver a grande maioria dos verbetes pertence a nomes derivados de verbos da primeira conjugação.

Tabela 36 - Fator vogal temática verbal: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
primeira conjugação	95	2.659.356	90	1.045.615	73	337286
o segunda conjugação	0,3	8.585	8	89.471	27	125767
o terceira conjugação	4,7	126.665	2	17.658	0	0
o TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Cabe destacar como os nomes deverbais terminados em *e* se desviam desse padrão geral, já que foram encontrados 27% de *types* com essa vogal final pertencendo à segunda conjugação. Isso pode ser resultado da alta frequência relativa das palavras *debate* (70.148 *tokens*) e *combate* (46.174 *tokens*), respectivamente, 15% e 10% dos nomes deverbais terminados em *e* (463.053 *tokens*).

5.2.2 Vogal temática

Em relação à variável vogal temática, na Tabela 37 temos o resultado inverso ao apresentado na tabela anterior, já que a alta frequência de alteração da vogal temática verbal se dá devido ao alto índice de *tokens* com vogal final *o* e *e*, associado ao fato de haver uma grande maioria de nomes deverbais derivados de verbos da primeira conjugação, o que replica uma tendência natural do português brasileiro para os nomes deverbais.

Tabela 37 - Fator vogal temática: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
alteração da VT	100	2.794.606	17	197.575	73	337.286
manutenção da VT	0	0	83	956.716	27	125767
		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

5.2.3 Complexidade morfológica

Em relação à variável complexidade morfológica, na Tabela 38 vemos que a tendência geral sem exceções é a presença de uma estrutura morfológica simples para todas as vogais finais, ou seja, sem acréscimo de afixos.

Tabela 38 - Fator Complexidade Morfológica: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
simples	90	2.511.319	81	928.907	77	358.290
complexa	10	283.287	19	225.384	33	104.763
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Porém percebe-se um decréscimo dessa tendência proporcional à frequência da vogal final, do maior índice para nomes terminados em *o* ao menor para nomes terminados em *e*, o que significa dizer que há mais probabilidade de haver nomes deverbais complexos morfologicamente terminando em *e* do que em *o* ou *a*.

5.2.4 Vogal média tônica

Em relação à variável vogal média tônica, a Tabela 39 indica que há uma tendência predominante de ausência de vogal média tônica, exceto para os nomes deverbais terminados em *a*.

Tabela 39 - Fator Vogal Média Tônica: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
ausência	71	1.997.607	43	492.941	81	373.176
presença	29	796.699	57	659.803	19	89.877
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Esse resultado pode estar refletindo a alta frequência do nome verbal *reforma* (111.907 *tokens*), com 10% de todos os nomes terminados em *a* (1.152.744 *tokens*).

5.2.5 Gênero da palavra

Conforme já observado anteriormente para os resultados dos *types*, a Tabela 40 indica que há uma correlação entre a tendência dos nomes deverbais com a vogal final *o* e a realização do gênero gramatical masculino.

Tabela 40 - Fator Gênero da Palavra: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
masculino	100	2.794.606	0	0	100	463.053
feminino	0	0	100	1.152.744	0	0
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

5.2.6 Número de sílabas

Em relação à variável número de sílabas, a Tabela 41 aponta para a frequência mais alta de nomes deverbais trissílabos: 86% para palavras terminadas em *o* (por exemplo, *re.tor.no*), 60% para palavras terminadas em *a* (por exemplo, *es.co.lha*), e 76% para palavras terminadas em *e* (por exemplo, *de.ba.te*).

Tabela 41 - Fator Número de Sílabas: Corpus Brasileiro

	o	Ocorrênci	a	Ocorrênci	e	Ocorrênci
	%	as	%	as	%	as
monossíla	0	0	0	0	0	0
bo						
dissílabo	5	147.118	39	450.967	18	81648
trissílabo	86	2.412.149	60	696.321	76	350327
polissabo	9	235.539	1	5.456	6	31078
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Ainda cabe destacar a alta frequência relativa de dissílabos terminados em *a* (39% dos *tokens* em *a*). Isso pode ser o resultado da alta frequência de *perda* (88.752) e *volta* (82.414), que somam 15% do total de *tokens* terminados em *a*.

5.2.7 Onset da última sílaba

Em relação à variável *onset* na última sílaba, a Tabela 42 indica que a tendência geral é de presença do *onset* nessa posição silábica, porém percebemos que há um leve desvio desse padrão para os nomes deverbais terminados em *o* (8% de *tokens* com ausência de *onset* na última sílaba), como por exemplo, em *bloqueio* (16.358 *tokens*, 3,5% do total de casos terminados em *e*).

Tabela 42 - Fator Onset da Última Sílab: Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
presença	92	2.577.305	99,9	1.151.575	100	463.053
ausência	8	217.301	0,1	1169	0	0
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Isso pode estar ocorrendo devido ao grande número de *tokens* com padrão -io na última sílaba para os nomes deverbais terminados em *o*.

5.2.8 Contexto fonológico precedente (modo de articulação)

Em relação à variável modo de articulação do contexto fonológico precedente, na Tabela 43 vemos que, para os nomes terminados em *a* e *e*, há uma frequência mais alta do contexto plosivo (por exemplo, *perda*, 88.752 *tokens*, e *debate*, 70.148 *tokens*).

Tabela 43 - Fator Contexto Fonológico Precedente (modo de articulação): Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
nasal	25	684.530	12	132.497	1	4.206
plosiva	21	578.107	51	590.409	90	415.642
fricativa	9	248.506	6	64.199	7	31.163
lateral	30	850.414	10	115.610	2	12.042
tepe	5	143.374	22	248.828	0	0
vogal	10	275.061	9	1.201	0	0
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

Por outro lado, para nomes deverbais terminados em *o*, foi o contexto lateral o mais relevante em termos de frequência, como, por exemplo, no caso de *desfile*, com 11.801 *tokens*.

5.2.9 Contexto fonológico precedente (ponto de articulação)

Conforme vemos na Tabela 44, em relação à variável ponto de articulação do contexto fonológico precedente, houve a frequência geral de alveolar para todas as vogais finais, que parece seguir uma tendência do léxico do português brasileiro para o contexto de *onset* (como nos casos de *confronto*, 22.755 tokens, *volta*, 82.414 tokens, e *ajuste*, 33.123 tokens).

Tabela 44 - Fator Contexto Fonológico Precedente (ponto de articulação): Corpus Brasileiro

	o %	Ocorrênci as	a %	Ocorrênci as	e %	Ocorrênci as
bilabial	5	138.479	11	123.942	2	7.562
labiodenta	1	38.583	0,1	1.853	0	0
l						
alveolar	45	1.244.473	67	768.913	78	362583
palatal	36	1.010.704	10	116.823	0,3	1536
velar	3	87.304	12	139.990	19	91.372
vogal alta	10	274.914	0,1	1.129	0	0
vogal	0	147	0	72	0	0
média						
vogal baixa	0	0	0	0	0	0
TOTAL		2.794.606		1.152.744		463.053

Fonte: Elaboração própria.

A seguir, na próxima seção, comparamos os resultados obtidos a partir do léxico dicionarizado e do léxico institucionalizado, em busca de convergências e divergências.

5.3 COMPARAÇÃO *TYPES* E *TOKENS*: LÉXICO DICIONARIZADO E LÉXICO INSTITUCIONALIZADO

Nesta seção comparamos os padrões observados na frequência dos *types* e dos *tokens*.

5.3.1 Vogal temática verbal

Para a variável vogal temática verbal, como podemos observar na Tabela 45, em termos gerais não há divergência quanto à alta frequência de nomes derivados de verbos pertencentes à primeira conjugação.

Tabela 45 - Comparação entre types e tokens: fator vogal temática verbal

	vogal o		vogal a		vogal e	
	types	tokens	types	tokens	types	tokens
primeira conjugação	98%	95%	94%	90%	95%	73%
segunda conjugação	1,4%	0,3%	5,4%	8%	5%	27%
terceira conjugação	0,6%	4,7%	0,6%	2%	0%	0%

Fonte: Elaboração própria.

Na verdade, o que chama bastante a atenção é a divergência entre a predominância no léxico dicionarizado desse tipo de nome (95%), no caso de vogal final *e*, que não se repete no léxico institucionalizado (73%). Conforme apontado no comentário sobre o padrão de frequência dessa variável no léxico institucionalizado, a frequência de 27% de nomes derivados de verbos da segunda conjugação está sendo encabeçada pelas palavras *debate* (70.148 *tokens*) e *combate* (46.174 *tokens*), respectivamente, 15% e 10% dos nomes deverbais terminados em *e* (463.053 *tokens*). O padrão ATE³⁹ pode ser um fator importante para a emergência da vogal final *e*, já que ATO não é um padrão muito frequente (*desacato*, 509 *tokens*; *desbarato*, 20 *tokens*; *trato*, 14.585 *tokens*); já o padrão ATA se manifesta apenas em *cata*, com 502 *tokens*. Para nomes deverbais terminados em *e*, ATE está presente em 10 *tokens*, (*abate*, 8.962 *tokens*; *combate*, 46.174 *tokens*; *debate*, 70.148 *tokens*; *desempate*, 1125 *tokens*; *desengate*, 18 *tokens*; *empate*, 12.878 *tokens*; *rebate*, 483 *tokens*; *remate*, 370 *tokens*; *resgate*, 15.933 *tokens* s; e *desbarate*, 4 *tokens*: um total de 156.091 *tokens*, 34% dos *tokens* de nomes deverbais terminados em *e*, e 14% dos *types*).

Esses resultados deixam clara a necessidade de uma abordagem exemplarista, já que podemos perceber como as sequências finais como ATE, que podem ser consideradas *types*

³⁹ Utilizamos a sequência com letras maiúsculas para distingui-las de categorias linguísticas como morfema ou sílaba.

fonológicos, mas não se referem a primitivos linguísticos como sílaba, morfema, pé silábico etc., são “salientes” para a aplicação de uma regra gramatical. Em outras palavras, a frequência de uma sequência sonora que não cabe em nenhuma categoria gramatical conhecida pode estar influenciando o processamento de uma regra na mente do falante simplesmente porque essa sequência emerge da alta frequência de palavras que contêm essa sequência final ATE.

5.3.2 Vogal temática

Para a variável vogal temática, conforme a Tabela 46, os resultados espelham a relação entre a variável conjugação verbal e a qualidade da vogal final dos nomes deverbais: como há uma alta incidência de verbos da primeira conjugação tanto em relação aos *types* quanto em relação aos *tokens*, e uma alta frequência de vogal final *o* e *e*, a maioria dos dados em *types* e *tokens* é de nomes que alteram a vogal temática da base verbal mais frequente.

Tabela 46 - Comparação entre types e tokens: fator vogal temática

	vogal a		vogal e		vogal o	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
alteração da VT	100%	100%	6%	17%	95%	73%
manutenção da VT	0%	0%	94%	83%	5%	27%

Fonte: Elaboração própria.

5.3.3 Complexidade morfológica

Para a variável complexidade morfológica, como vemos na Tabela 47, os nomes deverbais terminados em *o* e *a* apresentam resultados convergentes para *types* e *tokens*, com destaque para a elevação da frequência de *tokens* com estrutura morfológica simples para nomes terminado em *o* (90%), quando comparada a sua frequência de *types* (63%).

Tabela 47 - Comparação entre types e tokens: fator complexidade morfológica

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
simples	63%	90%	78%	81%	39%	77%
complexa	37%	10%	22%	19%	61%	33%

Fonte: Elaboração própria.

Já para nomes deverbais terminados em *e*, há divergência entre *types* e *tokens*: para *types*, há predomínio de formas complexas (61%) e para *tokens*, formas simples (77%).

5.3.4 Vogal média tônica

Para a variável vogal média tônica, a Tabela 48 mostra que há convergência entre a frequência de *types* e *tokens* para nomes terminados em *o* e *e*.

Tabela 48 - Comparação entre types e tokens: fator vogal média tônica

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
ausência	61%	71%	59%	43%	89%	81%
presença	39%	29%	41%	57%	11%	19%

Fonte: Elaboração própria.

Para os nomes terminados em *a*, a frequência de *types* indica a ausência de vogal média tônica como fator mais frequente (59%) e a frequência de *tokens* aponta a presença dessa vogal como o fator mais relevante (57%).

5.3.5 Gênero da palavra

Para a variável gênero da palavra, como pode ser observado na Tabela 49, há convergência entre *types* e *tokens* para todas as qualidades de vogal final.

Tabela 49 - Comparação entre types e tokens: fator gênero da palavra

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
masculino	100%	100%	0%	0%	100%	100%
feminino	0%	0%	100%	100%	0%	0%

Fonte: Elaboração própria.

5.3.6 Número de sílabas

Para a variável número de sílabas, conforme a Tabela 50, também há convergência entre *types* e *tokens* para todas as qualidades de vogal final.

Tabela 50 - Comparação entre types e tokens: fator número de sílabas

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
monossílabo	0%	0%	0%	0%	0%	0%
dissílabo	12%	5%	33%	39%	9%	18%
trissílabo	66%	86%	56%	60%	76%	76%
polissílabo	22%	9%	11%	1%	15%	6%

Fonte: Elaboração própria.

5.3.7 Onset da última sílaba

Como pode ser percebido na Tabela 53, para a variável *onset* da última sílaba, há convergência entre *types* e *tokens* para todas as qualidades de vogal final.

Tabela 51 - Comparação entre types e tokens: fator onset da última sílaba

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
presença	86%	92%	97%	99,9%	100%	100%
ausência	14%	8%	3%	0,1%	0%	0%

Fonte: Elaboração própria.

Aqui é interessante observar que, não só no caso dessa variável, a frequência de *tokens* parece potencializar as tendências observadas na frequência de *types*, aumentando a distância entre os fatores mais frequentes e os menos frequentes.

5.3.8 Contexto fonológico precedente (modo de articulação)

Para a variável modo de articulação do contexto fonológico precedente, como é exposto na Tabela 52, há convergência entre *types* e *tokens* para os nomes deverbais terminados em *a* e *e*.

Tabela 52 - Comparação entre types e tokens: fator contexto fonológico precedente (modo de articulação)

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
nasal	9%	25%	9%	12%	5%	1%
plosiva	29%	21%	44%	51%	76%	90%
fricativa	26%	9%	17%	6%	16%	7%
lateral	9%	30%	9%	10%	3%	2%
tepe	9%	5%	15%	22%	0%	0%
vogal	18%	10%	6%	9%	0%	0%

Fonte: Elaboração própria.

Para os nomes terminados em *o*, para *types*, o fator plosiva é o mais frequente (29%) e para *tokens*, o fator lateral (30%)

5.3.9 Contexto fonológico precedente (ponto de articulação)

Para a variável ponto de articulação do contexto fonológico precedente, na Tabela 53 vemos que há convergência entre *types* e *tokens* para todas as qualidades de vogal final.

Tabela 53 - Comparação entre types e tokens: fator contexto fonológico precedente (ponto de articulação)

	vogal o		vogal a		vogal e	
	292	2.794.606	128	1.152.744	74	463.053
	types	tokens	types	tokens	types	tokens
bilabial	4%	5%	6%	11%	9%	2%
labiodental	3%	1%	2%	0,1%	0%	0%
alveolar	44%	45%	52%	67%	55%	78%
palatal	18%	36%	13%	10%	4%	0,3%
velar	14%	3%	21%	12%	32%	19%
vogal alta	16%	10%	5%	0,1%	0%	0%
vogal média	0,3%	0%	0,8%	0%	0%	0%
vogal baixa	0%	0%	0%	0%	0%	0%

Fonte: Elaboração própria.

5.3.10 Convergências e divergências

Conforme podemos ver no Quadro 3, que resume o padrão de convergência e divergência entre a frequência de *types* e *tokens* das vogais finais *o*, *a* e *e*, houve divergência apenas em três das dez variáveis: complexidade morfológica, vogal média (tônica) e modo de articulação do contexto precedente.

Quadro 3 - Comparação entre *types* e *tokens*: convergências e divergências

	Convergência	Divergência
Conjugação verbal	<i>o; a; e</i>	
Vogal temática verbal	<i>o; a; e</i>	
Complexidade morfológica	<i>o; a</i>	<i>e</i>
Vogal média tônica	<i>o; e</i>	<i>a</i>
Gênero da palavra	<i>o; a; e</i>	
Número de sílabas	<i>o; a; e</i>	
Onset da última sílaba	<i>o; a; e</i>	
Contexto precedente (modo)	<i>a; e</i>	<i>o</i>
Contexto precedente (ponto)	<i>o; a; e</i>	

Fonte: Elaboração própria.

Os nomes deverbais terminados em *o* divergiram quanto ao modo de articulação do contexto precedente (*types*: **plosiva**; *tokens*: **lateral**); os nomes terminados em *a* divergiram quanto à vogal média tônica (*types*: **ausência**; *tokens*: **presença**); e os nomes terminados em *e* divergiram quanto à complexidade morfológica (*types*: **complexa**; *tokens*: **simples**).

5.3.11 Padrões de sequências finais

Nesta seção analisamos os padrões de sequências finais mais frequentes para *types* e *tokens*, de cada uma das três vogais dos nomes deverbais (para listagem completa de distribuição das palavras segundo a sequência, ver Apêndice A).

A seguir, na Tabela 54, temos a distribuição de ocorrência em *types* e *tokens* dos padrões de sequência final para os nomes terminados em *o*. Em *tokens*, as sequências mais frequentes são: -lho, -no e -do. Em *types*, -to, -eio e -jo.

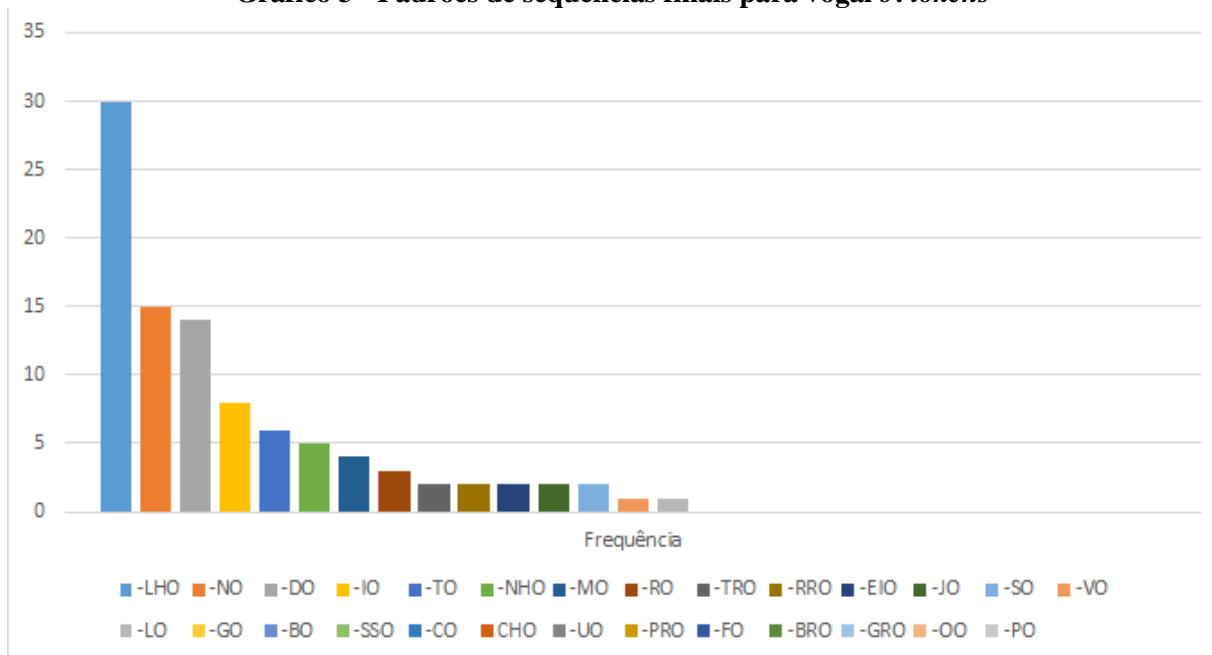
Tabela 54 - Padrões de sequências finais para vogal o: comparação entre types e tokens

tokens		%	types		%
-lho	819595	30	-to	31	12
-no	415024	15	-eio	28	11
-do	380398	14	-jo	23	9
-io	214371	8	-go	18	7
-to	168970	6	-ro	18	7
-nho	133829	5	-io	16	6
-mo	126507	4	-no	15	6
-ro	76306	3	-do	15	6
-tro	67503	2	-lo	13	5
-rro	61272	2	-cho	11	4
-eio	54720	2	-lho	9	3
-jo	51367	2	-rro	9	3
-so	47290	2	-sso	8	3
-vo	37249	1	-bo	6	2
-lo	30584	1	-vo	6	2
-go	14615	0.5	-co	5	2
-bo	11210	0.4	-nho	5	2
-sso	10404	0.3	-so	5	2
-co	6293	0.2	-fo	5	2
cho	6114	0.2	-bro	3	1
-uo	5815	0.2	-mo	3	1
-pro	3022	0.1	-tro	2	0.7
-fo	1339	0.04	-uo	2	0.7
-bro	678	0.02	-gro	1	0.4
-gro	443	0.01	-po	1	0.4
-oo	147	0.00	-pro	1	0.4
-po	4	0.00	-oo	1	0.4
TOTAL	2745069	100		260	100

Fonte: Elaboração própria.

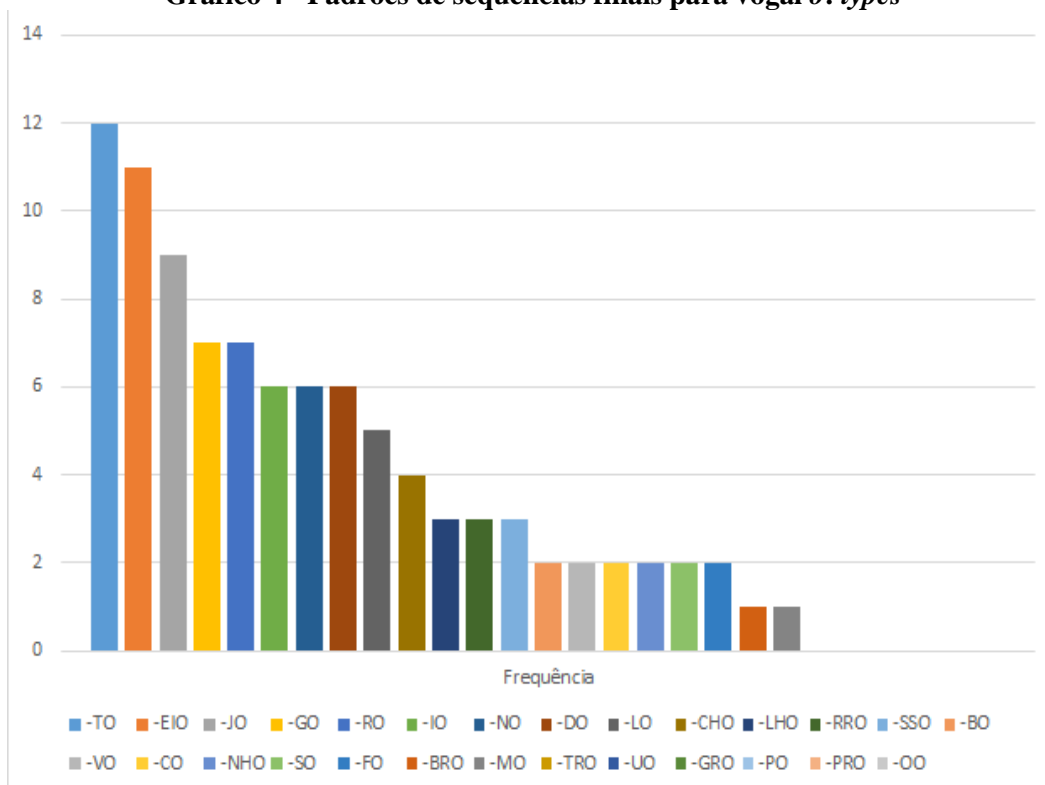
A seguir temos a informação fornecida pela Tabela 54 convertida para linguagem visual dos gráficos, para a distribuição em *tokens* e *types* (Gráficos 3 e 4).

Gráfico 3 - Padrões de seqüências finais para vogal o: tokens



Fonte: Elaboração própria.

Gráfico 4 - Padrões de seqüências finais para vogal o: types



Fonte: Elaboração própria.

A seguir, na Tabela 55, temos a distribuição de ocorrência em *types* e *tokens* dos padrões de sequência final para os nomes terminados em *a*. Em *types*, as sequências mais frequentes são: -da, -ta e -ma. Em *tokens*, -ta, -da e -ga.

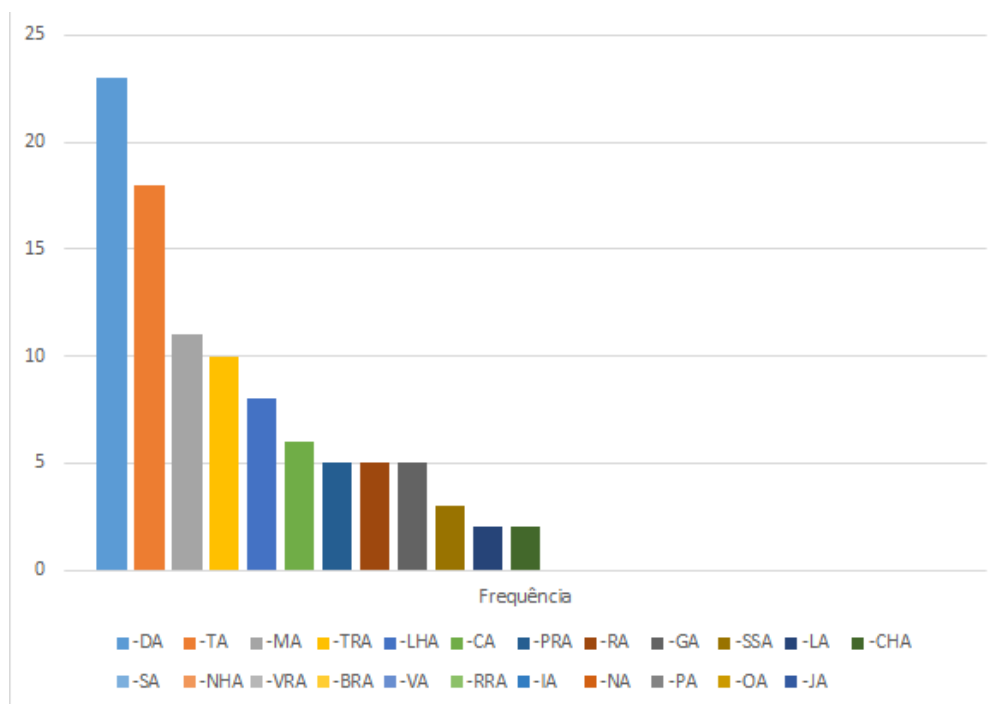
Tabela 55 - Padrões de sequências finais para vogal a: comparação entre types e tokens

tokens		%	types		%
-da	251842	23	-ta	13	12
-ta	196769	18	-da	12	11
-ma	120885	11	-ga	12	11
-tra	112184	10	-ra	9	8
-lha	85741	8	-ca	7	6
-ca	70130	6	-ia	6	6
-pra	55948	5	-sa	5	5
-ra	55228	5	-la	5	5
-ga	51819	5	-rra	5	5
-ssa	30472	3	-lha	5	5
-la	28276	2	-ma	4	4
-cha	21608	2	-va	3	3
-sa	8983	0.8	-nha	3	3
-nha	8435	0.7	-ssa	3	3
-vra	2860	0.2	-tra	2	2
-bra	2550	0.2	-ja	2	2
-va	1853	0.1	-na	2	2
-rra	1423	0.1	-pa	2	2
-ia	1129	0.1	-cha	1	1
-na	394	0.03	-bra	1	1
-pa	369	0.03	-pra	1	1
-oa	72	0.00	-vra	1	1
-ja	16	0.00	oa	1	1
TOTAL	1108986	100		105	100

Fonte: Elaboração própria.

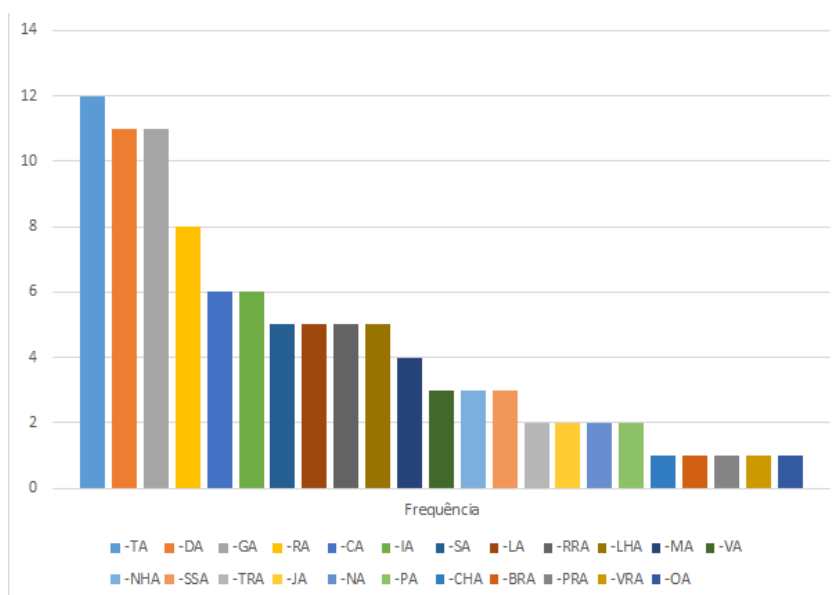
A seguir temos a informação fornecida pela Tabela 55 convertida para linguagem visual dos gráficos, para a distribuição em *tokens* e *types* (Gráficos 5 e 6).

Gráfico 5 - Padrões de sequências finais para vogal *a*: *tokens*



Fonte: Elaboração própria.

Gráfico 6 - Padrões de sequências finais para vogal *a*: *types*



Fonte: Elaboração própria.

Abaixo, na Tabela 56, temos a distribuição de ocorrência em *types* e *tokens* dos padrões de sequência final para os nomes terminados em *e*. Em *types*, as sequências mais frequentes são: -te, -que e -lhe; em *tokens*, -te, -que e -sse.

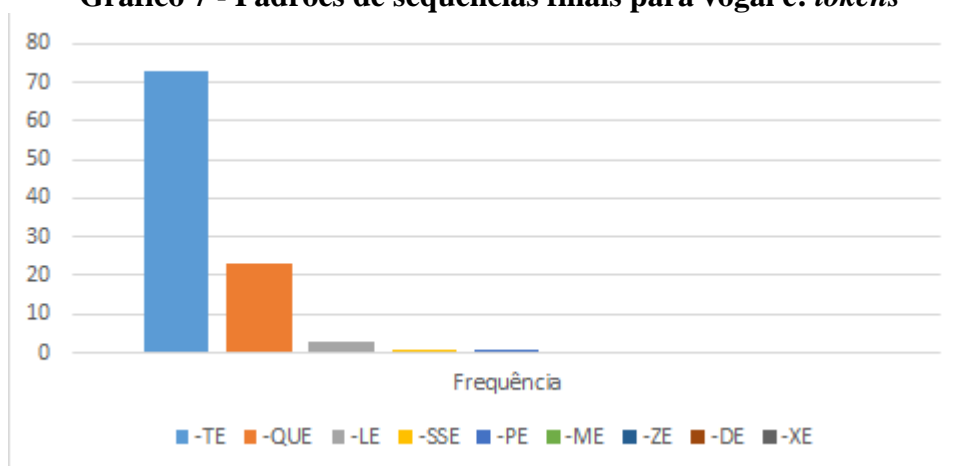
Tabela 56 - Padrões de seqüências finais para vogal e: comparação entre types e tokens

tokens		%	types		%
-te	297756	73	-te	20	41
-que	89718	23	-que	17	35
-le	11801	3	-sse	5	10
-sse	4259	1	-de	2	4
-pe	3209	1	-le	1	2
-me	1165	0.2	-me	1	2
-ze	803	0.1	-pe	1	2
-de	459	0.1	-xe	1	2
-xe	102	0.00	-ze	1	2
TOTAL	409272	100		49	100

Fonte: Elaboração própria.

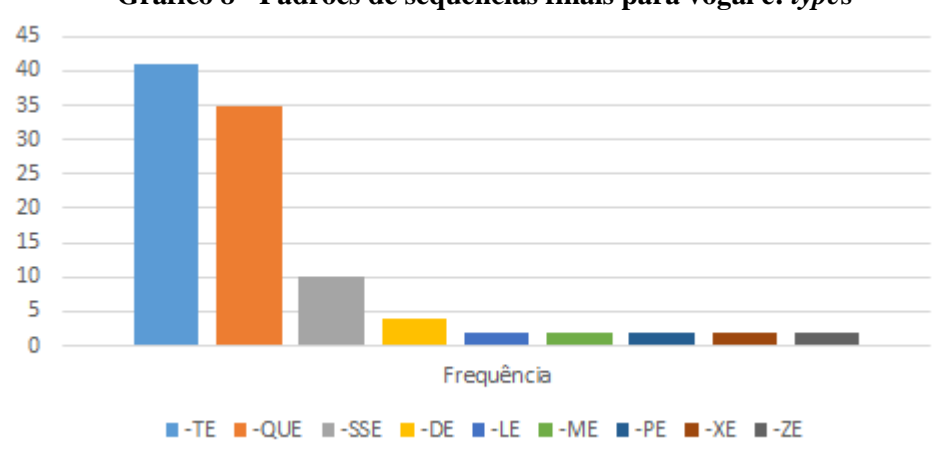
Abaixo temos a informação fornecida pela Tabela 56 convertida para linguagem visual dos gráficos, para a distribuição em *types* e *tokens* (Gráficos 7 e 8).

Gráfico 7 - Padrões de seqüências finais para vogal e: tokens



Fonte: Elaboração própria.

Gráfico 8 - Padrões de seqüências finais para vogal e: types



Fonte: Elaboração própria.

Como podemos observar no Quadro 4, o único grupo que não apresenta nenhuma convergência entre *types* e *tokens* para os padrões de sequência final se refere aos nomes terminados em *o*.

Quadro 4 - Comparação entre *tokens* e *types*: convergências e divergências nos padrões de sequências finais

	convergência	divergência
<i>o</i>		<i>types</i> : LHO, NO, DO <i>tokens</i> : TO, EIO, JO
<i>a</i>	<i>types</i> : DA, TA <i>tokens</i> : DA, TA	<i>types</i> : MA <i>tokens</i> : GA
<i>e</i>	<i>types</i> : TE, QUE <i>tokens</i> : TE, QUE	<i>types</i> : LE <i>tokens</i> : SSE

Fonte: Elaboração própria.

Conforme já discutido na seção 5.3.1, há novamente aqui uma alternativa oferecida pela Teoria dos Exemplos para explicar um fenômeno que está relacionado à memória e à emergência de padrões a partir da repetição, isto é, da frequência de uso de determinadas sequências sonoras que não pertencem às categorias linguísticas tradicionais de sílaba, morfema, pé silábico etc. Como já mencionado anteriormente, a frequência alta ou baixa dessas sequências sonoras pode estar intimamente relacionada à emergência da vogal final no processo morfológico da derivação regressiva. No entanto, essencialmente, o que isso implica para a descrição desse fenômeno? Isso implica que estamos diante de um fenômeno gramatical, ou seja, de processamento de regras correlacionando-se à padrões sonoros mais salientes (no sentido de frequência alta de ocorrência no uso da língua); desse modo, tornam-se menos distintas as fronteiras entre processamento e memória, entre representação e regra, como ditavam os preceitos dos modelos gerativos tradicionais.

5.4 EXPERIMENTO ONLINE COM PSEUDOPALAVRAS

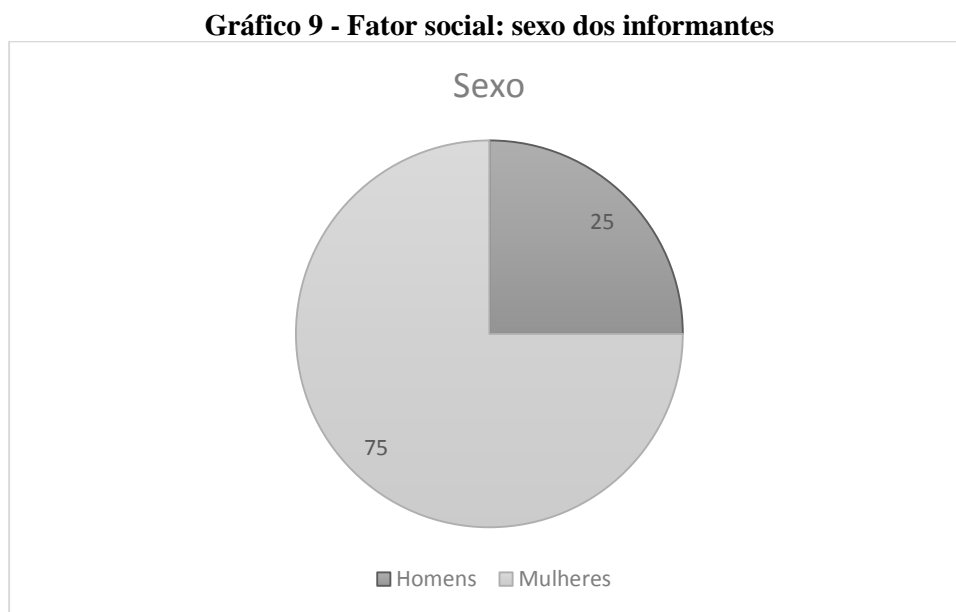
Nesta seção descrevemos os resultados do experimento online (Apêndice C), aplicado a 100 alunos de graduação, que tinha como propósito verificar qual das duas forças tem mais relevância para a produtividade da derivação regressiva sob a ótica da vogal final: **padrão feminino do léxico** ou **padrão masculino de formação da derivação regressiva**? Tomamos como hipótese, com base nos pressupostos da Teoria dos Exemplos, a relevância maior do padrão de formação da derivação regressiva para a emergência da vogal final. Todos os informantes tiveram acesso ao Termo de Consentimento Livre e Esclarecido (Apêndice B) e

concordaram em participar do experimento online. Antes de ser aplicado aos informantes, o experimento foi submetido ao Comitê de Ética e Pesquisa e aprovado pelo órgão (conforme Anexo A).

A seguir apresentamos primeiramente a estratificação de nossa amostra de 100 informantes em relação aos fatores sociais *sexo* e *faixa etária*, em seguida descrevemos e analisamos nossos resultados com base nas respostas dos informantes às 6 questões de múltipla escolha.

5.4.1 Fatores sociais

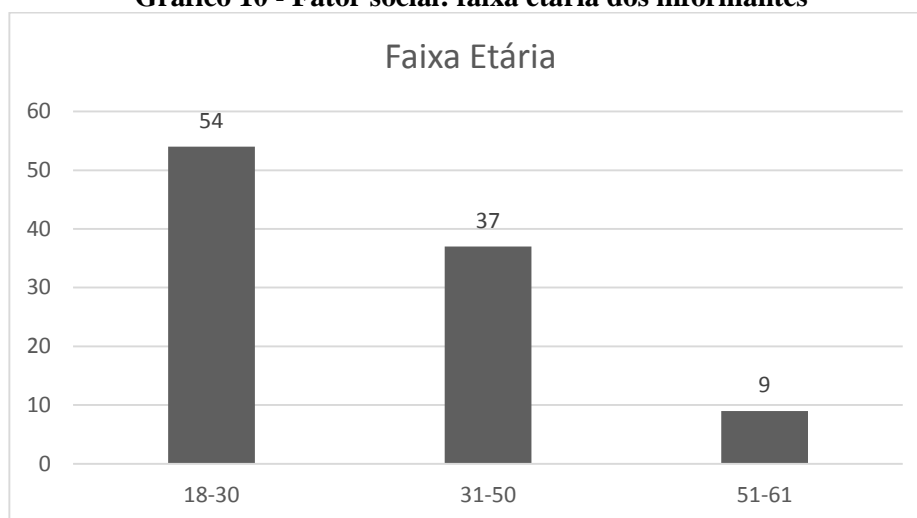
Abaixo, no Gráfico 9, temos a estratificação da amostra em relação ao fator sexo dos informantes. Como podemos observar, há o predomínio de informantes mulheres em nossa amostra.



Fonte: Elaboração própria.

Temos também, conforme podemos notar a seguir no Gráfico 10, a estratificação da amostra para o fator faixa etária dos informantes.

Gráfico 10 - Fator social: faixa etária dos informantes

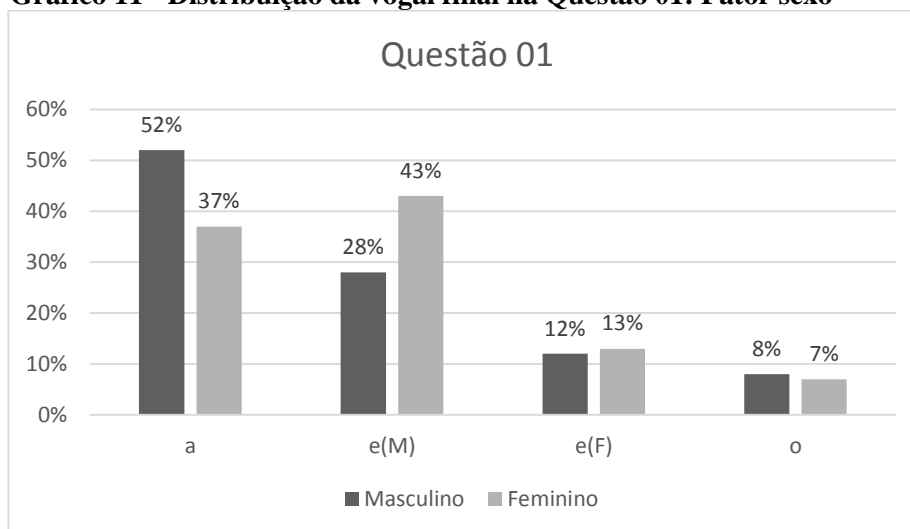


Fonte: Elaboração própria.

Os informantes estão distribuídos em 3 grupos etários: *jovens*, 18 a 30 anos (54%), *adultos*, 31 a 50 anos (37%), e *idosos*, 51 a 61 (9%).

Mesmo tendo em nossa amostra essas características de estratificação, não é nossa intenção realizar uma análise sociolinguística, por não considerarmos relevantes os fatores sociais para a produtividade da derivação regressiva. Apesar desse propósito, podemos perceber que há um padrão distinto no que se refere ao fator *sexo* dos informantes. Conforme vemos abaixo, no Gráfico 11, para a Questão 01, a maioria dos informantes do sexo masculino (52%) escolheu “sua maluta”, enquanto para informantes do sexo feminino, “seu malute” foi a escolha predominante (43%).

Gráfico 11 - Distribuição da vogal final na Questão 01: Fator sexo



Fonte: Elaboração própria.

Para as outras duas questões (02 e 03) que envolviam a escolha de uma pseudopalavra derivada do verbo no infinitivo, também houve divergência na forma escolhida: na Questão 02, “essa fetade” (40%) para sexo masculino, e “essa fetada” para sexo feminino; na Questão 03, “o mopito”, para homens, e “a mopita” para mulheres.

Acreditamos que essa discrepância entre os resultados quanto ao fator sexo se deve ao desequilíbrio na constituição da amostra segundo esse fator: 75 mulheres e 25 homens. Para podermos obter uma informação mais definitiva sobre o papel desse fator social, deveríamos ter uma amostra mais uniforme. Isso vale tanto para o fator sexo quanto para o fator faixa etária. A divisão em grupos etários também não atinge essa uniformidade necessária a uma análise sociolinguística (para uma visão geral da distribuição das respostas dos informantes ver Apêndice F).

Apesar dessa não uniformidade da amostra, para as questões que lidavam com a escolha de um pseudoverbo na forma infinitiva (04, 05 e 06), a opção predominante foi sempre a forma infinitiva da primeira conjugação (-ar), independentemente do fator social em foco.

É preciso salientar que nunca estipulamos como um propósito deste estudo realizar uma análise sociolinguística da derivação regressiva. A constituição de nossa amostra foi realizada de forma randômica; o único controle extralinguístico que existiu foi para que não houvesse estudantes do curso de Letras entre os informantes. Esse critério reduziu nossa amostra de 103 para 100 informantes, o que consideramos um número bastante significativo para verificar a produtividade da derivação regressiva em léxico potencial do PB.

A seguir apresentamos os resultados para as 6 questões de múltipla escolha do experimento online.

5.4.2 Seis questões de múltipla escolha

Nesta seção apresentamos em detalhes as 6 questões de múltipla escolha de nosso experimento, que foi aplicado a 100 informantes (todos cursando algum curso de graduação; exceto cursos de Letras, para não haver enviesamento nos resultados dessa amostra).

5.4.2.1 Análise dos resultados da questão 01

Diante da frase da questão 01, “João costuma MALUTER diariamente. Ele, contudo, precisa comer muito antes de”, os informantes tinham as seguintes opções para

completar a lacuna na frase: (i) sua maluta; (ii) seu malute; (iii) sua malute e (iv) seu maluto, em ordem aleatória (para a lista completa de respostas de cada informante, ver Apêndice E).

Conforme a Tabela 57, o pseudoverbo *maluter*, pertencente à segunda conjugação (grupo não produtivo no PB), apresentou como forma mais frequente a opção “sua maluta”, seguida de “seu malute”.

Tabela 57 - Respostas dos informantes à Questão 01

Pseudopalavras	Frequência
sua maluta	41
seu malute	39
sua malute	13
seu maluto	7

Fonte: Elaboração própria.

Temos acima uma forma feminina e uma forma masculina com frequências muito próximas, o que pode ser um indicativo de que formas verbais pertencentes às conjugações menos produtivas do PB podem apresentar resultados distintos das formas pertencentes à conjugação mais produtiva, ou seja, formas verbais com vogal temática *a*.

5.4.2.2 Análise dos resultados da questão 02

Diante do texto da questão 02, “Pedro tenta FETADIR sempre antes de dormir. Seu amigo recomendou para melhorar sua concentração.”, os informantes tinham as seguintes opções para completar a lacuna na frase: (i) essa fetada; (ii) essa fetade; (iii) esse fetado e (iv) esse fetade, em ordem aleatória.

Conforme a Tabela 58, o pseudoverbo *fetadir*, pertencente à terceira conjugação (grupo menos produtivo no PB), apresentou como forma mais frequente “essa fetada”, seguido de “essa fetade”.

Tabela 58 - Respostas dos informantes à Questão 02

Pseudopalavras	Frequência
essa fetada	40
essa fetade	27
esse fetado	17
esse fetade	16

Fonte: Elaboração própria.

Temos acima duas formas femininas como as mais frequentes, o que corrobora a direção apontada pelos resultados da questão 01: para as formas verbais pertencentes aos grupos menos

produtivos (segunda e terceira conjugações), a forma preferida pelos informantes é a forma feminina terminada em vogal *a*.

5.4.2.3 Análise dos resultados da questão 03

Diante do texto da questão 03, “Maria não tem medo de MOPITAR enquanto dirige. Mas já foi avisada que pode ser um crime grave de trânsito.”, os informantes tinham as seguintes opções para completar a lacuna na frase: (i) o mopito; (ii) o mopite; (iii) a mopita e (iv) a mopite, em ordem aleatória.

Segundo vemos na Tabela 59, o pseudoverbo *mopitar*, pertencente à primeira conjugação (grupo verbal mais produtivo no PB), apresentou como forma mais frequente “o mopito”, seguido de “o mopite”.

Tabela 59 - Respostas dos informantes à Questão 03

Pseudopalavras	Frequência
o mopito	42
o mopite	30
a mopita	18
a mopite	10

Fonte: Elaboração própria.

Observamos acima uma tendência oposta à observada na questão 01 e na questão 02: para a forma verbal com a vogal temática mais produtiva, os informantes selecionaram a forma derivada masculina terminada na vogal *o*, que é a mais típica para esse fenômeno de derivação regressiva.

Com base nos resultados acima, poderíamos considerar como uma tendência relevante para a formação de formas derivadas de verbos via processo da derivação regressiva a correlação entre vogal temática mais e menos produtiva e a emergência das três vogais finais nominais, *o*, *a* e *e*.

Passamos agora à segunda parte dos resultados, que diz respeito aos casos de seleção de uma forma infinitiva a partir de sua contraparte derivada via derivação regressiva.

5.4.2.4 Análise dos resultados da questão 04

Diante do texto da questão 04, “Nos aposentos da casa, devido ao excesso de umidade, somente com ZENETO mantemos a casa limpa. Se esquecermos de , sofreremos com

alergias o ano inteiro.”, os informantes tinham as seguintes opções para completar a lacuna na frase: (i) *zenetar*; (ii) *zeneter*; e (iii) *zenetir*, em ordem aleatória.

Conforme a Tabela 60, para a pseudopalavra *zeneto*, a forma mais frequente foi o infinitivo na primeira conjugação “*zenetar*”, com a maioria quase absoluta da preferência dos informantes.

Tabela 60 - Respostas dos informantes à Questão 04

Pseudopalavras	Frequência
zenetar	83
zeneter	10
zenetir	7

Fonte: Elaboração própria.

Este é um resultado bastante esperado quando consideramos a produtividade no PB das formas verbais da primeira conjugação.

5.4.2.5 Análise dos resultados da questão 05

Diante do texto da questão 04, “Cada vez mais comum nas democracias do mundo, optar pela BOPLIDA é uma prática bastante popular entre o público jovem. Se não puderem se sentem antiquados.”, os informantes tinham as seguintes opções para completar a lacuna na frase: (i) *boplidar*; (ii) *boplidir*; e (iii) *boplider*, em ordem aleatória.

Assim como observado na questão 05, temos, na Tabela 61, para a pseudopalavra *boplida*, a forma mais frequente “*boplidar*”, pertencente à primeira conjugação.

Tabela 61 - Respostas dos informantes à Questão 05

Pseudopalavras	Frequência
boplidar	72
boplidir	22
boplider	6

Fonte: Elaboração própria.

Aqui também é interessante notar o número mais significativo de “*boplidir*”, da terceira conjugação, a menos produtiva no PB, o que pode ser resultado de um processo de harmonia vocálica entre a vogal -i- do radical e a vogal temática -i- (*boplidir*). O processo de harmonia vocálica pode ser um fator relevante em competição para emergência da vogal final via processo de derivação regressiva.

Conforme Schwindt (2007, p. 398), a harmonia vocálica verbal, com base em experimentos com pseudopalavras, apresenta resultados bastante consistentes em termos de transparência dessa regra para verbos da terceira conjugação, como é o caso da opção “boplidir”:

[...] harmonia na terceira conjugação se mostrou transparente pelos resultados do Segundo teste. Diante de formas flexionadas de pseudoverbos, os informantes relacionaram as vogais /i,u/, presentes na raiz, à vogal temática -i em quase 80% dos casos, tanto no presente do indicativo quanto no presente do subjuntivo. Isso sugere que a Harmonia Vocálica Verbal, embora apresente uma baixa produtividade na língua, é um processo transparente para os falantes do PB. (SCHWINDT, 2007, p. 398)⁴⁰

Os resultados apontados por Schwindt (2007) podem explicar a alta ocorrência de casos de “boplidir”, um verbo da terceira conjugação, que não pertence à conjugação mais produtiva do PB, já que poderíamos estar diante da necessidade por parte dos falantes de manutenção da harmonia entre a vogal -i- do radical e a vogal temática verbal -i-.

5.4.2.6 Análise dos resultados da questão 06

Diante do texto da questão 04, “Para garantir a privacidade das mensagens eletrônicas, executa-se o FITUQUE antes de seu envio. Ao logar no aplicativo, é só uma vez para garantir que as mensagens sejam protegidas de invasores.”, os informantes tinham as seguintes opções para completar a lacuna na frase: (i) fitucar; (ii) fituquir; e (iii) fituquer, em ordem aleatória.

Tabela 62 - Respostas dos informantes à Questão 06

Pseudopalavras	Frequência
fitucar	81
fituquir	16
fituquer	3

Fonte: Elaboração própria.

A questão 06 replica o padrão observado na questão 05, a forma “fitucar”, da primeira conjugação aparece como a mais frequente (também observado na questão 04), e a forma

⁴⁰ [...] harmony in the third conjugation was shown to be transparent by the results of the second test. Faced with inflected pseudo-verbs, the informants related the vowels /i, u/, present in the root, to the theme vowel -i in almost 80% of the cases, both in the present indicative and in the present subjunctive. This suggests that VH, despite its low productivity in the language, is a transparent process for BP speakers. (SCHWINDT, 2007, p. 398)

“fituquir” é a segunda mais frequente, o que parece indicar novamente a influência do processo de harmonia vocálica na seleção da forma infinitiva.

Com base nos resultados a seguir (Quadro 5), a forma da primeira conjugação é sempre a mais frequente, o que é esperado quando consideramos o fator produtividade no PB, já que as formas verbais da primeira conjugação são sempre as mais frequentes e as mais produtivas no PB. O resultado que chama mais atenção, porém, é a seleção como segunda opção da forma infinitiva de terceira conjugação, que é geralmente a menos frequente no PB. Quando analisamos mais de perto essas formas, descobrimos como a harmonia vocálica pode ter um papel relevante para a emergência da vogal final envolvida no processo de derivação regressiva, conforme mencionado na seção anterior.

Quadro 5 - Resultado geral do experimento online: pseudopalavras mais frequentes

Lista de pseudopalavras mais frequentes
mal <u>u</u> ta
fetada <u>a</u>
mopit <u>o</u>
zenet <u>a</u> r
boplid <u>a</u> r
fituc <u>a</u> r

Fonte: Elaboração própria.

A Tabela 63 cruza os resultados de cada uma das questões que envolviam a escolha de um nome derivado a partir da forma infinitiva de um pseudoverbo (Questões 1, 2 e 3) com as respostas dadas pelos informantes para cada uma das questões que envolviam a escolha de um pseudoverbo no infinitivo a partir de um nome deverbal (Questões 4, 5 e 6). O que esse quadro revela é a consistência na relação entre a escolha das vogais finais para os nomes deverbais e a escolha da conjugação do pseudoverbo: como a frequência de verbos da primeira conjugação apresentou um índice elevado de ocorrência, independentemente da vogal final escolhida pelos informantes, não houve em nenhum momento a alteração desse fato. Isso pode indicar que para os falantes não há necessidade de preservar a correlação entre vogal temática do verbo que serve de base para a derivação e a vogal temática do nome que é derivado desse verbo. Como veremos mais adiante quando comparamos o resultado desse experimento aos resultados do léxico dicionarizado (dicionários *Aurélio* e *Houaiss*) e léxico institucionalizado (*Corpus Brasileiro*), essa tendência na produção de novas formas parece corroborar o que se observa na frequência de *types* e *tokens*.

Tabela 63 - Cruzamento de respostas dos informantes: forma derivada x forma infinitiva

	zenet-			boplid-			fituc-		
seu maluto	ar	er	-ir	ar	er	-ir	ar	er	-ir
	3	0	0	3	0	0	3	0	0
sua maluta	ar	er	-ir	ar	er	-ir	ar	er	-ir
	31	8	2	29	3	9	34	1	6
seu malute	ar	er	-ir	ar	er	-ir	ar	er	-ir
	34	2	3	26	2	11	31	1	7
sua malute	ar	er	-ir	ar	er	-ir	ar	er	-ir
	12	0	0	10	0	2	8	1	3
esse fetado	ar	er	-ir	ar	er	-ir	ar	er	-ir
	16	0	1	10	1	6	14	1	2
essa fetada	ar	er	-ir	ar	er	-ir	ar	er	-ir
	30	7	3	28	4	8	33	0	7
esse fetade	ar	er	-ir	ar	er	-ir	ar	er	-ir
	14	2	0	13	1	2	12	2	2
essa fetade	ar	er	-ir	ar	er	-ir	ar	er	-ir
	23	1	3	21	0	6	22	0	5
o mopito	ar	er	-ir	ar	er	-ir	ar	er	-ir
	36	4	2	32	4	6	30	1	11
a mopita	ar	er	-ir	ar	er	-ir	ar	er	-ir
	15	2	1	15	0	3	16	0	2
o mopite	ar	er	-ir	ar	er	-ir	ar	er	-ir
	15	2	1	15	0	3	16	0	2
a mopite	ar	er	-ir	ar	er	-ir	ar	er	-ir
	24	2	4	17	1	12	27	1	2

Fonte: Elaboração própria.

Diferentemente do observado acima na Tabela 63, abaixo (Tabela 64) observamos que há algumas divergências entre a tendência geral na escolha do pseudonome deverbais e no que se observa quando cruzamos esse resultado com cada uma das respostas dos informantes para as Questões 1, 2 e 3. Em metade das respostas dadas pelos informantes, podemos encontrar inconsistências com o padrão geral: por exemplo, os informantes que escolheram “seu maluto” como resposta para a Questão 1, em sua maioria escolheram “essa fetade” como resposta para a Questão 2, o que contraria a tendência geral, “essa fetada”. Para as respostas que se referem

às escolhas com frequência mais alta (“sua maluta”, “essa fetada” e “o mopito”), não foram encontradas inconsistências no cruzamento dos resultados, o que parece indicar que as inconsistências podem estar relacionadas à frequência mais baixa dos itens em questão.

Tabela 64 - Cruzamento das respostas dos informantes: formas derivadas x formas derivadas
(continua)

	fetad-				mopit-			
seu maluto	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	2	2	0	3	3	1	1	2
sua maluta	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	6	16	6	13	19	7	11	4
seu malute	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	9	13	9	8	15	8	14	2
sua malute	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	0	9	1	3	5	2	4	2
	malut-				mopit-			
esse fetado	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	2	6	9	0	7	2	7	1
essa fetada	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	2	16	13	9	18	7	12	3
esse fetade	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	0	6	9	1	8	2	4	2

Tabela 64 - Cruzamento das respostas dos informantes: formas derivadas x formas derivadas (conclusão)

	malut-				mopit-			
essa fetade	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	3	13	8	3	9	7	7	4
	malut-				fetad-			
o mopito	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	3	19	15	5	7	18	8	9
a mopita	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	1	7	8	2	2	7	2	7
o mopite	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	1	11	14	4	7	12	4	7
a mopite	o	a	e (m)	e (f)	o	a	e (m)	e (f)
	2	4	2	2	1	3	2	4

6.5 COMPARAÇÃO ENTRE LÉXICO DICIONARIZADO, LÉXICO INSTITUCIONALIZADO E LÉXICO POTENCIAL

Nesta seção comparamos os resultados observados no léxico dicionarizado (dicionários Aurélio e Houaiss), no léxico institucionalizado (*Corpus Brasileiro*) e no léxico potencial (experimento com pseudopalavras). A metodologia utilizada para comparação dos resultados, primeiro léxico dicionarizado e léxico institucionalizado, e depois as convergências dessa primeira análise comparadas aos resultados do experimento online, justifica-se pela necessidade de inicialmente buscar correlações entre *types* (dicionário) e *tokens* (*Corpus*

Brasileiro), para em seguida corroborar ou não essas convergências com os resultados do experimento online.

Conforme vemos no Quadro 6, em relação ao fator “complexidade morfológica”, para a emergência da vogal final *e* parece ser relevante a ocorrência dos prefixos *des-* e *re-*. Quanto ao fator “vogal média tônica”, os nomes deverbais terminados em *a* parecem preferir o contexto em que ocorre a vogal média na sílaba tônica. No que se refere ao “modo de articulação do contexto precedente”, percebemos que a vogal final *o* pode ser sensível ao segmento lateral. O padrão sonoro final é consistente para *a* e *e*; para *o* somente se observaram divergências.

Quadro 6 - Comparação: Dicionário x Corpus Brasileiro x Experimento online

	RESULTADOS
DICIONÁRIO x <i>CORPUS BRASILEIRO</i>	fatores mais relevantes para a seleção da vogal final: complexidade morfológica vogal média tônica contexto precedente (modo) padrão sonoro final
EXPERIMENTO	fatores mais produtivos: vogal final: <i>a</i> e <i>o</i> conjugação verbal: 1ª conjugação

Fonte: Elaboração própria.

Quando comparamos esses resultados com aqueles encontrados no experimento com pseudopalavras, esperamos uma frequência maior de *o* para todas as questões que envolviam a escolha de um nome deverbal, mas não foi isso que os resultados apontaram: das três questões, duas tiveram como resposta mais frequente a forma terminada em *a*, o que segue o padrão geral dos substantivos no português brasileiro (SCHWINDT, 2018). Assim, conforme os resultados do experimento online, vemos duas forças em conflito na emergência da vogal final dos nomes deverbais: **(i) a tendência geral dos substantivos em emergir a vogal final *a*; (ii) a tendência específica dos nomes deverbais formados por derivação regressiva em emergir a vogal final *o*.**

As pseudopalavras dessas três questões apresentavam os contextos precedentes /t/ e /d/: *malut-*, *fetad-* e *mopit-*. Porém, esses contextos não parecem ter condicionado as respostas dos informantes, já que com contexto /t/ tivemos *maluta* e *mopito* como os itens mais frequentes. O domínio da forma feminina terminada em *a* parece indicar que a correlação entre a vogal temática do verbo e do nome sofre uma pressão para preservar sua qualidade vocálica. Enquanto na correlação entre o nome e o verbo, o critério mais relevante é a produtividade da primeira conjugação. Assim, a vogal temática verbal parece ser o fator mais relevante tanto na relação

entre “verbo > nome derivado” quanto na relação “nome derivado > verbo”. Sem a ocorrência de prefixos, a vogal final *e* não se destacou; ao contrário de *a*, que apresentou frequência alta mesmo sem a ocorrência do contexto vogal média tônica, ou *o*, sem a ocorrência do contexto precedente modo lateral. Pode-se concluir, então, que há uma competição entre as formas terminadas em *o* e *a*, em que dois fatores parecem medir forças para a emergência da vogal final: a produtividade da primeira conjugação e o predomínio em geral de nomes terminados em *o*. Ao neutralizar, no léxico potencial, os possíveis contextos influenciadores de cada uma das vogais finais, *o*, *a* e *e*, observados no léxico dicionarizado e no léxico institucionalizado, percebe-se um favorecimento da vogal final *a* na produção de formas potenciais no português brasileiro, o que fortalece a percepção de que há um papel mais relevante da tendência geral dos substantivos (SCHWINDT, 2018) para o léxico potencial.

6 CONSIDERAÇÕES FINAIS

Neste estudo, defendemos a hipótese de que **os padrões produtivos do português brasileiro sinalizam que a derivação regressiva é um processo especial de nominalização, que toma o radical da palavra como base da derivação, acrescentando uma vogal final que é influenciada por fatores linguísticos.** Para isso, analisamos a produtividade da derivação regressiva sob a ótica da vogal final em três amostras de léxico: o léxico dicionarizado, o léxico institucionalizado e o léxico potencial. Entendemos a produtividade em dois âmbitos: a **disponibilidade** da derivação regressiva no léxico dicionarizado no PB, a partir de dados obtidos do Dicionário Eletrônico *Aurélio* (e corroborados pelo Dicionário Eletrônico *Houaiss*); e a **rentabilidade** desse processo morfológico no léxico institucionalizado, por meio da frequência lexical dos dados no *Corpus Brasileiro*, e no léxico potencial, por meio de um experimento online com pseudopalavras, aplicado a 100 alunos de ensino superior.

Nosso objetivo geral era **investigar a produtividade da derivação regressiva sob a ótica da vogal final.** Para isso, organizamos nossa análise a partir de 3 objetivos específicos:

1. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico dicionarizado (Dicionários Eletrônicos Houaiss e Aurélio);
2. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico institucionalizado (*Corpus Brasileiro*);
3. Investigar a produtividade da derivação regressiva sob a ótica da vogal final em amostra do léxico potencial (experimento online com pseudopalavras).

Orientando-nos pela hipótese e objetivos apontados acima, obtivemos os resultados que seguem.

1. Léxico dicionarizado e léxico institucionalizado

Na etapa de levantamento de dados do léxico dicionarizado, foram obtidos 759 dados (após a reanálise de dados excluídos por pertencerem a categorias especiais como regionalismos, formas com *homofonia acidental* e formas com alternância na vogal final), que foram posteriormente codificados segundo fatores fonológicos e morfológicos a fim de verificar sua influência nos padrões de produtividade da derivação regressiva em relação à emergência da vogal final, que pode ser qualquer uma das três vogais *o*, *a* e *e*. Para cada etapa metodológica buscamos responder uma questão:

(a) Há uma relação entre a vogal temática dos verbos derivantes e a vogal temática dos nomes derivados?

Nossos resultados apontam para uma possível relação entre nomes deverbais terminados em *o* e verbos da primeira conjugação (-ar). Há uma alta frequência de nomes terminados em *o* para os casos de derivação regressiva, o que não reflete a tendência geral para os substantivos no português brasileiro, predomínio de formas terminadas em *a* (SCHWINDT, 2018). Quanto aos verbos da primeira conjugação, são os mais produtivos no português brasileiro, e dessa confluência de altas frequências de conjugação do verbo e de vogal final dos nomes deverbais parece emergir essa relação.

Um resultado referente à frequência divergente para a ocorrência de verbos da segunda conjugação para vogal final *e* aponta para a necessidade de uma abordagem exemplarista. Nos *types* 95% dos casos de vogal final *e* referem-se a verbos da primeira conjugação; por outro lado, no *Corpus* esse número baixa para 73%. Isso se deve ao fato de que, no *Corpus Brasileiro*, os 27% de nomes derivados de verbos da segunda conjugação estão sendo encabeçados pelas palavras *debate* (70.148 *tokens*) e *combate* (46.174 *tokens*), respectivamente, 15% e 10% dos nomes deverbais terminados em *e* (463.053 *tokens*). Para nomes deverbais terminados em *e*, ATE está presente em 10 *types*, (*abate*, 8.962 *tokens*; *combate*, 46.174 *tokens*; *debate*, 70.148 *tokens*; *desempate*, 1125 *tokens*; *desengate*, 18 *tokens*; *empate*, 12.878 *tokens*; *rebate*, 483 *tokens*; *remate*, 370 *tokens*; *resgate*, 15.933 *tokens*; e *desbarate*, 4 *tokens*: um total de 156.091 *tokens*, 34% dos *tokens* de nomes deverbais terminados em *e*, e 14% dos *types*). Esses resultados sugerem a demanda por uma abordagem exemplarista, já que podemos perceber como as sequências finais como ATE (que podem ser consideradas *types* fonológicos, mas não se referem a primitivos linguísticos como sílaba, morfema, pé silábico etc.) são “salientes” para a aplicação de uma regra gramatical.

(b) O que é mais frequente: a alternância da vogal temática da base verbal, como em atacAr/ataquE, ou a preservação da vogal, como em falAr/falA?

A questão acima está implicada nos resultados apresentados para a questão (a). Se parece haver uma correlação entre verbos da primeira conjugação e a emergência da vogal final *o*, logo verificamos uma tendência à alternância da vogal temática verbal no nome derivado, preferencialmente a alternância *a/o*. Ao mesmo tempo, também temos o caso da alta frequência do padrão final ATE, que parece indicar uma tendência à manutenção da vogal temática da segunda conjugação *e*.

(c) Qual é o tipo preferido de base para aplicação da regra de derivação regressiva: palavras simples, palavras prefixadas ou palavras sufixadas?

Quanto à complexidade morfológica dos nomes deverbais, observamos que nomes terminados em *o* e *a* mostram uma tendência a emergir em palavra simples; porém nomes deverbais terminados em *e*, para *types*, há predomínio de formas complexas (61%) e para *tokens*, formas simples (77%). Isso deve ao fato de haver um grande número de nomes prefixados terminados em *e* que estão disponíveis no léxico dicionarizado (60% dos 74 casos de nomes deverbais terminados em *e*) e que têm uma frequência baixa no *Corpus Brasileiro*.

(d) Formas verbais que apresentam alternância vocálica nas formas rizotônicas podem ser consideradas um indício de que as formas regressivas não tomam como base a forma flexionada, mas sim o morfema de radical como base para a derivação?

Quanto à presença de vogal média rizotônica, encontramos poucos casos que justificassem a afirmação de que as formas regressivas tomam como base o radical e não a forma flexionada (nos casos em que há contexto para a harmonia vocálica, ou seja verbos de segunda e terceira conjugação, cf. Schwindt 2007). Somente nos nomes terminados em *a*, observamos uma frequência de *types* significativa com vogal média rizotônica, 58% dos casos; mas, como em sua maioria absoluta referem-se a verbos da primeira conjugação, não há contexto para a harmonia vocálica.

(e) Os nomes formados via derivação regressiva seguem a tendência dos nomes em geral no português brasileiro quanto ao gênero mais frequente?

Nossos dados apontam para uma tendência divergente em relação à tendência geral dos nomes no português brasileiro. Enquanto os nomes em geral apresentam formas femininas terminadas em *a*, o padrão para nomes deverbais é a forma masculina terminada em *o*: 66% de *types* no léxico dicionarizado e 59% de *tokens* no léxico institucionalizado, no léxico institucionalizado.

(f) Palavras com duas ou três sílabas são um alvo mais frequente da derivação regressiva? ou o número de sílabas não é um fator relevante para o fenômeno da derivação regressiva?

O padrão geral do português brasileiro apresenta predomínio de polissílabos (VIARO, GUIMARÃES-FILHO, 2007). Nossos resultados parecem apontar para uma tendência diversa: encontramos em nossos dados uma frequência maior de trissílabos. Porém, como essa distribuição geral se refere ao léxico como um todo e não somente à classe dos substantivos,

não é possível afirmar se esse resultado do fator número de sílabas reflete um padrão específico dos nomes deverbais ou um padrão mais geral da classe dos substantivos.

(g) Há influência da presença de um segmento que precede a vogal final dos nomes deverbais? Será que ter um *onset* favorece o processo de derivação?

Em relação à variável *onset* da última sílaba, a maioria quase absoluta dos nomes deverbais apresenta *onset* nessa posição silábica. Porém, assim como no caso do número de sílabas, não é possível afirmar se esse resultado do fator número de sílabas reflete um padrão específico dos nomes deverbais ou um padrão mais geral da classe dos substantivos.

(h) A frequência de determinado modo de articulação do segmento precedente pode estar condicionando a emergência da vogal final?

No que se refere ao modo de articulação do segmento precedente, para os nomes terminados em *a* e *e*, há uma frequência mais alta do contexto plosivo. Por outro lado, para nomes deverbais terminados em *o*, para *types*, o fator plosiva é o mais frequente (29%) e para *tokens*, o fator lateral (30%). Esses segmentos são abordados de um modo exemplarista quando investigamos em seguida quais são os padrões sonoros mais frequentes para nomes deverbais, o que pode nos indicar que os resultados díspares para as vogais finais pode ser fruto das sequências sonoras mais salientes para os falantes e não apenas de um segmento precedente.

(i) a frequência de determinado ponto de articulação do segmento precedente pode estar condicionando a emergência da vogal final?

Para a variável ponto de articulação do contexto fonológico precedente, alveolar é o ponto de articulação mais frequente para todas as vogais finais. Assim como mencionado na questão (h), quando abordados de um modo exemplarista, verificando quais são os padrões sonoros mais frequentes para nomes deverbais, podemos ter uma indicação de como sequências sonoras mais salientes para os falantes estão influenciando a emergência da vogal final, e não apenas considerando a relevância de um segmento precedente.

2. Análise dos padrões sonoros finais mais frequentes

Para ilustrar a força da sequência de segmentos finais para emergência da vogal final apresentamos novamente o Quadro 4, que apresenta as convergências e divergências entre *types* e *tokens*.

Quadro 4 - Comparação entre *types* e *tokens*: convergências e divergências nos padrões de sequências finais

	convergência	divergência
<i>o</i>		<i>types</i> : LHO, NO, DO <i>tokens</i> : TO, EIO, JO
<i>a</i>	<i>types</i> : DA, TA <i>tokens</i> : DA, TA	<i>types</i> : MA <i>tokens</i> : GA
<i>e</i>	<i>types</i> : TE, QUE <i>tokens</i> : TE, QUE	<i>types</i> : LE <i>tokens</i> : SSE

Fonte: Elaboração própria.

A Teoria dos Exemplos parece ser o modelo mais adequado para explicar as convergências observadas acima para as vogais finais *a* e *e*, pois esse resultado parece indicar que estamos diante de um fenômeno associado à memória e à emergência de padrões a partir da repetição, isto é, da frequência de uso de determinadas sequências sonoras que não pertencem às categorias linguísticas tradicionais de sílaba, morfema, pé silábico etc. A frequência alta ou baixa dessas sequências sonoras pode estar intimamente relacionada à emergência da vogal final no processo morfológico da derivação regressiva. Isso implica uma diminuição da separação categórica entre processamento e memória, entre representação e regra, como ditavam os preceitos dos modelos gerativos tradicionais.

3. A comparação entre *types* e *tokens*

A força de uma regra gramatical pode ser medida pela convergência entre seu padrão no léxico dicionarizado e sua frequência em um corpus de referência que reflita os padrões de uso dos falantes. Em nosso estudo, percebemos que, quando comparados os resultados encontrados no Dicionário Eletrônico *Aurélio* e no *Corpus Brasileiro*, há convergência nos seguintes aspectos com alta frequência:

- proximidade dos resultados de distribuição da vogal final dos *types* do léxico e do *corpus*, mesmo com a redução de 35%, quando excluídos os nomes não encontrados no *corpus*;
- predomínio de formas derivadas de verbos da primeira conjugação;
- maioria de formas com atribuição de gênero masculino;
- alta frequência de nomes trissílabos;
- maioria de palavras com presença de *onset* na última sílaba;
- frequência mais alta do ponto de articulação alveolar do contexto precedente.

4. Léxico potencial

Diante de 6 questões de múltipla escolha, aplicadas online para 100 informantes, obtivemos os seguintes resultados, já apresentados anteriormente no Quadro 5:

Quadro 5 - Resultado geral do experimento online: pseudopalavras mais frequentes

Lista de pseudopalavras mais frequentes
maluta
fetada
mopito
zenetar
boplidar
fitucar

Fonte: Elaboração própria.

Acima podemos ver que há um número maior de nomes deverbiais terminados em *a*, seguindo a tendência geral para os nomes no português brasileiro, e divergindo do padrão previsto para nomes formados por derivação regressiva, que é o predomínio de palavras terminadas em *o*. Há duas forças que podem estar influenciando a emergência da vogal final *a*: uma delas é a já mencionada tendência geral dos nomes no português brasileiro; a outra, é a sequência final TA e DA, que se apresentam com alta frequência no *Corpus Brasileiro*. Podemos aventar até mesmo a possibilidade de haver uma confluência desses dois fatores aqui para a emergência da vogal final. No que se refere à maioria absoluta de pseudoverbos na forma infinitiva da primeira conjugação (-ar), não houve surpresa, já que essa conjugação é a mais produtiva no português brasileiro, mesmo que haja por parte de formas da segunda conjugação uma tendência a manter a vogal temática verbal na forma derivada.

5. A comparação entre léxico dicionarizado, léxico institucionalizado e léxico potencial

Ao compararmos os resultados de *types*, *tokens* e experimento com pseudopalavras, encontramos os seguintes resultados:

- convergência para os resultados do léxico e do *corpus* em relação ao predomínio de formas da primeira conjugação;
- divergência em relação ao predomínio da vogal final *o*, que ficou atrás da vogal final *a*.

A convergência refere-se a um fato que não tem relação direta com a derivação regressiva, mas está implicado na alternância entre vogal temática verbal e vogal final do nome deverbal: a primeira conjugação é a mais produtiva no português brasileiro. A força desse fator

parece ser muito relevante quando o associamos à tendência dos nomes deverbais em terminarem com a vogal *o*. Porém, quando os informantes selecionaram o pseudônimo deverbal mais adequado diante das opções disponíveis no português brasileiro, houve uma forte preferência pela vogal final *a*, o que segue a tendência geral para os nomes no português, e que também pode estar seguindo uma tendência observada em *types* e *tokens* para as sequências sonoras finais TA e DA.

Resumidamente, diante dos resultados para todas essas 5 etapas de nosso estudo, podemos afirmar que há uma relação entre fatores linguísticos e a emergência da vogal final, mesmo que se perceba uma tensão entre fatores associados ao léxico dicionarizado, ao léxico institucionalizado e à produtividade da relação entre vogal temática verbal e vogal final dos nomes derivados. Em *types* e *tokens*, percebemos a relevância das sequências finais, que parecem ter maior importância, quando analisamos o resultado do experimento online, do que a tendência observada dos nomes deverbais a emergir a vogal final *o*. Ainda cabe investigar mais a fundo o quanto um experimento mais amplo, com mais questões, abrangendo mais padrões sonoros finais, pode revelar sobre o papel da frequência dessas sequências finais para a emergência da vogal final.

Ademais, vemos como o padrão geral de distribuição da vogal final na derivação regressiva no léxico institucionalizado se reflete na forma como se organizam as formas deverbais conforme a vogal final no léxico dicionarizado (Gráfico 2). Esse fato confirma nossa escolha pela abordagem teórica da Teoria de Exemplos, que assume os padrões de uso como ponto de partida para a emergência das regras linguísticas que estruturam o conhecimento do falante.

Por fim, ainda existem questões que não puderam ser contempladas em nossa pesquisa sobre a produtividade da derivação regressiva e que merecem aprofundamento em estudos posteriores:

1. A necessidade de verificar de forma mais exaustiva o léxico potencial. Para isso, por exemplo, poderíamos analisar no léxico institucionalizado se os falantes produzem formas via derivação regressiva que já são formadas por outro processo de nominalização, como a sufixação por *-mento* e *-ção*.

2. A elaboração de novos experimentos que contemplem outros processos de nominalização. Como em nosso experimento somente foram apresentados casos de derivação

regressiva, há necessidade de construir novos instrumentos que acrescentem pseudopalavras formadas por sufixação em -ção e -mento.

3. A investigação da influência de fatores sociais na derivação regressiva. Como não era o propósito de nossa análise apresentar uma discussão sobre os aspectos sociolinguísticos da derivação regressiva, não nos aprofundamos nos fatores *sexo* e *faixa etária* que constituíram nossa amostra do experimento online. Como já mencionado no Capítulo 3,

A derivação regressiva com apoio na vogal *a* só ocorre em formações tradicionais, ao contrário da baseada em *o*, utilizada em formações recentes de caráter gíriático[...]. Formações em *e* são também produtivas, sobretudo em verbos prefixados com *re-* e *des-*, nos quais a derivação regressiva é relativamente comum. (BASILIO, 2011, p. 38-39)

Os termos “formações tradicionais” e “formações recentes” podem estar apontando para uma tensão muito relevante para os estudos sociolinguísticos entre “novo” e “velho”, entre formas consideradas mais eruditas e formas inovadoras.

REFERÊNCIAS

- ALCÂNTARA, Cíntia da Costa. As classes formais do português brasileiro. **Letras de Hoje**, Porto Alegre, v. 45, n. 1, p. 5-15, jan./mar. 2010.
- ANDERSON, S. R. **A-morphous morphology**. Cambridge: Cambridge University Press, 1992.
- ARONOFF, Mark. **Word formation in Generative Grammar**. Cambridge: MIT Press, 1976.
- BASILIO, Margarida. **Estruturas Lexicais do Português**. Petrópolis: Vozes, 1980.
- BASILIO, Margarida. **Teoria Lexical**. São Paulo: Ática, 1987.
- BASILIO, Margarida. **Formação e classes de palavras no português do Brasil**. São Paulo: Contexto, 2011.
- BAAYEN, R. H. Quantitative aspects of morphological productivity. In Booij, G. E., and Marle, J. van (Eds), **Yearbook of Morphology 1991**, Kluwer Academic Publishers, Dordrecht, 109-149, 1992.
- BAUER, Laurie. **Introducing linguistic morphology**. Edinburgh: Edinburgh University Press, 1988.
- BAUER, Laurie. **Morphological Productivity**. Cambridge: Cambridge University Press, 2001.
- BECHARA, Evanildo. **Moderna Gramática Portuguesa**. Rio de Janeiro: Nova Fronteira, 2009.
- BYBEE, Joan. **Phonology and Language Use**. Cambridge: Cambridge University Press, 2001.
- BYBEE, Joan. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. **Language Variation and Change** 14, p. 261-290, 2002.
- BYBEE, Joan. **Language, Usage and Cognition**. Cambridge: Cambridge University Press, 2010.
- CAMARA Jr., Joaquim M. **Problemas de linguística descritiva**. Petrópolis: Vozes, 1969.
- CAMARA Jr., Joaquim M. **Estrutura da Língua Portuguesa**. Petrópolis: Vozes, 1970.
- CORBIN, D. **Morphologie dérivationnelle et structuration du lexique**. Tubinga: Max Niemeyer Verlag, 1987.
- CUNHA, Celso; CINTRA, Lindley. **Nova Gramática do Português Contemporâneo**. Rio de Janeiro: Lexicon, 2017.

CRISTÓFARO SILVA, Thaís. Descartando fonemas: a representação mental na fonologia de uso. In: Dermeval da Hora; Gisella Collischonn. (Org.). **Teoria Lingüística: Fonologia e Outros Temas**. João Pessoa: Universidade Federal da Paraíba, p. 200-231. 2003.

DYLAN, Bob. The Times They Are A-Changin'. In: **Bob Dylan. The Lyrics: 1961-2012**. New York: Simon & Schuster, 2016.

FERREIRA, Aurélio Buarque de Holanda. **Dicionário Eletrônico Aurélio**. Rio de Janeiro: Editora Nova Fronteira e Lexikon Informática, 2001. Versão 5.0.40.

FREITAS, Margareth de Souza. **Derivação regressiva na perspectiva do modelo Dressler: algumas considerações**. 1990. 101f. Dissertação (Mestrado em Linguística) –Universidade Federal do Paraná, Curitiba.

GARCIA, Rafael. ONG aponta alta de 15% na taxa de desmate da Amazônia e confirma tendência. **Folha de São Paulo**. São Paulo, 16 ago. 2019. Disponível em: <<https://www1.folha.uol.com.br/ambiente/2019/08/desmatamento-cresce-15-nos-ultimos-12-meses-diz-ong.shtml>>. Acesso em: 16 ago. 2019.

HOUAISS, Antônio; VILLAR, Mauro de Salles (Dir.) **Dicionário Eletrônico Houaiss**. São Paulo: Objetiva, 2009. Versão 3.0.

LEE, Seug-Hwa. **Morfologia e fonologia lexical do português do Brasil**. Tese (Doutorado) – Universidade de Campinas, Campinas, 1995.

LIPKA, Leonhard. **An Outline of English Lexicology Lexical Structure, Word Semantics, and Word-Formation**. Tübingen: Gunter Narr, 1992.

LOBATO, Lúcia. A derivação regressiva em português: conceituação e tratamento gerativo. In: PEREIRA, Cilene da Cunha; PEREIRA, Paulo Roberto Dias (Orgs.). **Miscelânea de estudos linguísticos, filológicos e literários in memoriam Celso Cunha**. Rio de Janeiro: Nova Fronteira. p. 205-230. 1995.

MORENO, Cláudio. **Morfologia nominal do português: um estudo de fonologia lexical**. Tese (Doutorado) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1997.

PIERRENHUBERT, J. Exemplar Dynamics: Word frequency, lenition and contrast. In BYBEE, J.; HOPPER, P. (Ed.). **Frequency and the emergence of linguistic structure**. Amsterdam: Benjamins, 2001. p. 123-157.

QUADROS, Emanuel Souza de. **A estrutura e o uso da parassíntese no português**. 2009. Trabalho de Conclusão de Curso (Graduação em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre. 54 p.

RESENDE, Maurício Sartori. **Derivação regressiva e construções com verbos leves: um estudo sobre aspecto lexical**. 2016. 124f. Dissertação (Mestrado em Linguística) – Universidade Federal do Paraná, Curitiba.

RESENDE, Maurício Sartori. Reconciliando propostas lexicalistas para a derivação regressiva. **Macabéa** – Revista Eletrônica do Netlli, Crato, v. 5, n. 2, p. 53-61, jul.-dez. 2016.

ROCHA, Luiz Carlos de Assis. **Estruturas morfológicas do português**. Belo Horizonte: Ed. UFMG, 1998.

ROCHA LIMA, Carlos Henrique da. **Gramática normativa da língua portuguesa**. 49.ed. Rio de Janeiro: José Olympio, 2011.

RODRIGUES, Alexandra Soares. Capítulo 1 - Introdução. In: RIO-TORTO, Graça. et al. (coords.): **Gramática derivacional do Português**. Coimbra: Imprensa da Universidade de Coimbra., 2013a.

RODRIGUES, Alexandra Soares. Capítulo 2.4 - Nomes deverbais. In: RIO-TORTO, Graça. et al. (coords.): **Gramática derivacional do Português**. Coimbra: Imprensa da Universidade de Coimbra., 2013b.

SAID ALI, Manoel. **Gramática histórica da língua portuguesa**. São Paulo: Melhoramentos, 1921.

SARDINHA, Tony Berber; MOREIRA FILHO, José Lopes; ALAMBERT, Eliane. **Corpus Brasileiro**. São Paulo: CEPRI, LAEL, CNPq, Fapesp, PUCSP, 2010.

SARDINHA, Tony Berber. Lingüística de Corpus: histórico e problemática. Revista **DELTA Documentação de Estudos em Linguística Teórica e Aplicada**, v. 16, n. 2, p. 323-367, 2000.

SCHWINDT, L. C. Exponência de gênero e classe temática em português brasileiro. **Revista DELTA. Documentação de Estudos em Linguística Teórica e Aplicada**, v. 34, p. 745-768, 2018.

SCHWINDT, L. C. Paradigmatic correspondences in the brazilian portuguese verbal vowel System. **Acta Linguistica Hungarica**, v. 54, p. 393-40, 2007.

SCHWINDT, L. C.. Predizibilidade da marcação de gênero em substantivos no português brasileiro. In: Danniell Carvalho; Dorothy Brito. (Org.). **Gênero e língua(gem): formas e usos**. Salvador: Editora da UFBA, 2020, v. 1, p. 279-294.

SPENCER, Andrew. **Morphological Theory**. Cambridge: Basil Blackwell, 1991.

VIARO, Mário; GUIMARÃES-FILHO, Zwinglio. Análise quantitativa da frequência dos fonemas e estruturas silábicas portuguesas. **Estudos Linguísticos**, vol 36, p. 27 / 36, 2007.

**APÊNDICE A – PADRÕES NAS TERMINAÇÕES DOS SUBSTANTIVOS
DEVERBAIS COLETADOS DO DICIONÁRIO ELETRÔNICO AURÉLIO**

Padrões nas terminações dos substantivos deverbiais coletados do Dicionário Eletrônico Aurélio

Palavras terminadas em *o*

EIO	JO	TO	GO	RO
1. acarreio	1. adejo	1. acalanto	1. abodego	1. amparo
2. alardeio	2. almejo	2. acarreto	2. abrigo	2. anteparo
3. anseio	3. amojo	3. acato	3. achego	3. aparo
4. aperreio	4. anojo	4. acerto	4. adergo	4. apuro
5. arqueio	5. arejo	5. acosto	5. afago	5. chilro
6. balanceio	6. arquejo	6. acrescento	6. afogo	6. choro
7. bambaleio	7. arranjo	7. agito	7. alago	7. conjuro
8. bamboleio	8. arrojo	8. aperto	8. apego	8. descarro
9. barateio	9. bocejo	9. apresto	9. arrenego	9. desdoiro
10. bloqueio	10. bodejo	10. apronto	10. barrego	10. desdouro
11. boleio	11. boquejo	11. arrasto	11. berrego	11. desespero
12. bombardeio	12. bordejo	12. arrebatado	12. carrego	12. desvairo
13. borboleteio	13. bracejo	13. arrebitado	13. castigo	13. disparo
14. borneio	14. cacarejo	14. arreto	14. choramingo	14. encaro
15. cabeceio	15. cantarejo	15. arrote	15. conchego	15. enforo
16. cabeio	16. chamejo	16. assalto	16. derrengo	16. escancarar
17. cambaleio	17. cotejo	17. assesto	17. desafogo	17. esconjuro
18. campeio	18. dardejo	18. atesto	18. desassossego	18. exagero
19. careio	19. desajoujo	19. calafeto	19. descarrego	19. exaspero
20. carteio	20. desanojo	20. carreto	20. desembargo	20. namoro
21. ceceio	21. despejo	21. conforto	21. desemprego	21. paio
22. cerceio	22. doidejo	22. confronto	22. desengasgo	22. preparo
23. chalreio	23. doudejo	23. conserto	23. desengonço	23. regiro
24. chilreio	24. farejo	24. desacato	24. emprego	24. reparo
25. chuleio	25. farfalhejo	25. desaperto	25. encarrego	25. respiro
26. coleio	26. festejo	26. desaponto	26. engasgo	26. restauro
27. costeio	27. forcejo	27. desbarato	27. enxugo	27. retiro
28. custeio	28. gaguejo	28. desconto	28. escorrego	28. reverbero
29. desbloqueio	29. gargarejo	29. desencanto	29. estrafejo	29. tiro
30. desenfreio	30. gorgolejo	30. desfruto	30. estrago	
31. desenleio	31. gracejo	31. desgasto	31. expurgo	
32. desnorteio	32. lacrimejo	32. empreito	32. fumego	
33. enleio	33. latejo	33. encanto	33. fungo	
34. escaramuceio	34. manejo	34. enxerto	34. gingo	
35. espeloteio	35. mercadejo	35. espanto	35. ofego	
36. esperneio	36. ornejo	36. experimento	36. pago	
37. floreio	37. padejo	37. grito	37. refugio	
38. folheio	38. partejo	38. pipito	38. reprego	
39. fraseio	39. pastejo	39. protesto	39. resfôlego	
40. galanteio	40. pestanejo	40. quebranto	40. resmungo	
41. galeio	41. rastejo	41. reconto	41. respingo	
42. galreio	42. relampejo	42. rejeito	42. rogo	
43. garganteio	43. revoejo	43. repto	43. sossego	
44. garreio	44. rojo	44. ressalto	44. trasfego	
45. gateio	45. sacolejo	45. retorno	45. xingo	
46. gazeio	46. trinculejo	46. rusto		
47. geio	47. vaquejo	47. sobressalto		
48. gorjeio	48. vasquejo	48. sustento		
49. granjeio	49. verdejo	49. transporte		
50. maneio	50. vicejo	50. trato		
51. manuseio	51. voejo			
52. meneio				
53. norteio				
54. ondeio				
55. orneio				
56. palanqueio				
57. paleio				
58. papagueio				
59. pareio				

60. passeio 61. ponteio 62. prazenteio 63. rabeio 64. rasqueteio 65. rateio 66. recenseio 67. refreio 68. regateio 69. regorjeio 70. remeneio 71. resmoneio 72. revoluteio 73. rodeio 74. salteio 75. sapateio 76. saqueio 77. saracoteio 78. sorteio 79. tartamudeio 80. tartareio 81. titubeio 82. tonteio 83. torneio 84. toureio 85. trauteio 86. tuteio 87. veraneio 88. verdeio 89. volteio 90. vozeio 91. zangarreio 92. zanguizarreio				
---	--	--	--	--

LO	IO	NO	LHO	DO
1. abalo 2. açulo 3. acúmulo 4. adolo 5. apelo 6. atropelo 7. cochilo 8. consolo 9. degelo 10. degolo 11. desmantelo 12. dissímulo 13. embalo 14. empapelo 15. engabelo 16. engabelo 17. engrambelo 18. esfacelo 19. pipilo 20. pulo 21. regelo 22. rejúbilo 23. restilo 24. rutilo 25. tamborilo 26. trescalo 27. ululo 28. vacilo	1. aboio 2. alívio 3. apoio 4. arrepio 5. assobio 6. avio 7. balbucio 8. caio 9. câmbio 10. contracâmbio 11. desafio 12. desapoio 13. desmaio 14. envio 15. desperdício 16. extravio 17. mio 18. pipio 19. recâmbio 20. reenvio 21. rodopio 22. rufio 23. sítio 24. tenetio 25. transvio 26. tresvario	1. abandono 2. abono 3. aceno 4. afano 5. desabono 6. descortino 7. desempenho 8. desengano 9. desensino 10. detono 11. empeno 12. empino 13. ensino 14. governo 15. grasno 16. rebusno 17. rechino 18. refino 19. ressono 20. retorno 21. suborno 22. tino 23. transtorno 24. treino 25. trino 26. trono	1. agasalho 2. arrolho 3. arrulho 4. atalho 5. atulho 6. descoalho 7. desembrulho 8. desencalho 9. desentulho 10. desenxovalho 11. desfolho 12. desrefolho 13. encalho 14. enxovalho 15. esbulho 16. farfalho 17. mergulho 18. ralho 19. recolho 20. remolho 21. tafulho 22. talho 23. trabalho 24. tresmalho	1. abordo 2. acordo 3. aguardo 4. arremedo 5. comando 6. desagrado 7. desando 8. desbordo 9. descuido 10. desenfardo 11. desenredo 12. desmando 13. discordo 14. enredo 15. escaldo 16. nado 17. olvido 18. remedo 19. resguardo 20. respaldo 21. retardo 22. transbordo 23. trasbordo

RRO	ÇO	CO	CHO	OO
1. acorro	1. aboço	1. arranco	1. acocho	1. acoo
2. aferro	2. abraço	2. atravanco	2. agacho	2. atroo
3. aforro	3. alcanço	3. cerco	3. atochó	3. desentoo
4. aterro	4. ameaço	4. choco	4. bochecho	4. enjoo
5. berro	5. arregaço	5. coco	5. cochicho	5. esboroo
6. desaferro	6. avanço	6. confisco	6. cuincho	6. escachoo
7. desaterro	7. calço	7. decalco	7. desabrocho	7. escoo
8. desemperro	8. derraço	8. derranco	8. despachó	8. reboó
9. desforro	9. desemeço	9. descerco	9. desmancho	9. ressoo
10. desgarro	10. desengonço	10. emborco	10. despacho	10. retroo
11. desterro	11. destroço	11. fabrico	11. empachó	11. sobrevoó
12. emperro	12. emboço	12. farisco	12. escrachó	12. troo
13. empurro	13. ercalço	13. mexerico	13. esculacho	13. voó
14. encerro	14. entrelaço	14. pipoco	14. relincho	
15. enterro	15. escaço	15. rebusco	15. trincho	
16. erro	16. esforço	16. remanisco		
17. esbarro	17. lanço	17. risco		
18. esborro	18. rechaço	18. saçarico		
19. esparro	19. reforço	19. troco		
20. espirro	20. revição	20. trovisco		
21. socorro	21. tropeço			
22. urro	22. descarrego			

BO	NHO	VO	XO	BRO
1. afobo	1. acanho	1. agravo	1. amouxo	1. assombro
2. arroubo	2. alinhó	2. conchavo	2. coaxo	2. cobro
3. desgabo	3. apanho	3. cultivo	3. desleixo	3. desdobro
4. esculhambo	4. arreganho	4. desconchavo	4. empuxo	4. dessoçobro
5. estrabo	5. assanho	5. desestorvo	5. encaixo	5. recobro
6. menoscabo	6. descaminho	6. mascavo	6. estrouxo	6. requebro
7. rebimbo	7. desempenho	7. relevo	7. rebaixo	7. ressumbro
8. retumbo	8. despenho	8. silvo	8. remelexo	8. soçobro
9. roubo	9. escanganho	9. sorvo	9. repuxo	
10. tombo	10. solinho			
11. zumbo				

FO	SSO	TRO	UO	GRO	DRO
1. abafo	1. acosso	1. desemplastro	1. acuo	1. malogro	1. desmedro
2. barrufo	2. arremesso	2. desencontro	2. amuo	2. reintegro	2. ladro
3. bufo	3. confesso	3. deslastro	3. recuo		
4. desabafo	4. endosso	4. deslustró			
5. desarrufo	5. repasso	5. encontro			
6. esborrifo	6. traspasso				

Palavras terminadas em *a*

TA	CA	GA	DA	RA
1. afronta	1. arranca	1. achega	1. aguarda	1. conjura
2. ajunta	2. cabroca	2. chega	2. ajuda	2. cora
3. ausculta	3. cabruca	3. delonga	3. contenda	3. dependura
4. cata	4. coca	4. desfadiga	4. demanda	4. destêmpera
5. chapota	5. descasca	5. desobriga	5. emenda	5. espera

6. consulta 7. conta 8. corta 9. cresta 10. desafronta 11. desfruta 12. desleita 13. desmoita 14. desmouta 15. disputa 16. empreita 17. escuta 18. espreita 19. experimenta 20. fita 21. grita 22. janta 23. permuta 24. pernoita 25. pernouta 26. reconquista 27. reestampa 28. repergunta 29. requesta 30. revista 31. solta 32. visita 33. volta	6. despesca 7. destroca 8. ensaca 9. inculca 10. machuca 11. marca 12. merca 13. perca 14. pesca 15. prática 16. prédica 17. rebusca 18. réplica 19. risca 20. ronca 21. saca 22. sapeca 23. seca 24. súplica 25. tasca 26. tréplica 27. trisca 28. troca	6. entrega 7. esfrega 8. larga 9. liga 10. nega 11. obriga 12. outorga 13. paga 14. pega 15. pesga 16. prolonga 17. rega 18. respiga 19. rezinga 20. rocega 21. salga 22. sega 23. sonega 24. sossega 25. trasfega 26. vaga 27. voga	6. engorda 7. escalda 8. foda 9. guarda 10. lida 11. monda 12. muda 13. poda 14. reemenda 15. referenda 16. revenda 17. tolda 18. veda	6. gira 7. joeira 8. jura 9. melhora 10. mira 11. paquera 12. pendura 13. penhora 14. piora 15. procura 16. recaptura
--	---	---	--	---

MA	LHA	VA	LA	RRA
1. chama 2. derrama 3. desmama 4. desrama 5. doma 6. embroma 7. lástima 8. prema 9. queima 10. reforma 11. requeima 12. subestima 13. superstima 14. toma	1. colha 2. debulha 3. desfolha 4. desolha 5. encolha 6. escolha 7. esgalha 8. malha 9. molha 10. pilha 11. recolha 12. talha 13. trilha	1. ceva 2. conserva 3. desova 4. encrava 5. escova 6. esquivia 7. estorva 8. recova 9. recurva 10. seva 11. sova 12. trava	1. chancela 2. charla 3. degola 4. despela 5. enrola 6. escabela 7. fala 8. fila 9. parla 10. tala 11. vela	1. agarra 2. desforra 3. embirra 4. empurra 5. encerra 6. ferra 7. garra 8. surra 9. torra

ÇA	NA	PA	CHA	JA
1. avança 2. caça 3. coça 4. derriça 5. destrinça 6. espinça 7. orça 8. troça	1. capina 2. descorna 3. encarna 4. esgana 5. guina 6. tisna	1. capa 2. carpa 3. empa 4. limpa 5. reestampa	1. chucha 2. desmancha 3. escancha 4. marcha 5. pedincha	1. beija 2. manja 3. peja 4. peleja 5. vareja

CIA	SA	NHA	XA	BRA
1. denúncia 2. despronúncia	1. escusa 2. repisa	1. apanha 2. barganha	1. deixa 2. queixa	1. quebra 2. sobra

3. pronúncia 4. renúncia	3. sopresa 4. tosa	3. ordenha 4. resenha	3. rebaixa	3. soçobra
-----------------------------	-----------------------	--------------------------	------------	------------

ZA	UA	BA	TRA	SSA
1. encruza 2. reza	1. desjeja 2. recua	1. derruba 2. surriba	1. amostra 2. mostra	1. acossa 2. remessa

Palavras terminadas em *e*

TE	QUE	CE	DE	XE
1. abate 2. aceite 3. ajuste 4. alevante 5. aparte 6. arraste 7. arremate 8. bote 9. combate 10. corte 11. debate 12. decote 13. desajuste 14. desate 15. desbarate 16. desbaste 17. desbote 18. descante 19. descarte 20. desempate 21. desengate 22. desfrute 23. desgaste 24. desmonte 25. despiste 26. desquite 27. empaste 28. empate 29. encarte 30. encoste 31. enfarte 32. engaste 33. esgote 34. implante 35. invite 36. levante 37. pernoite 38. pernoute	1. açambarque 2. arranque 3. ataque 4. calque 5. debique 6. decalque 7. derranque 8. descasque 9. desembarque 10. desfalque 11. despenque 12. despique 13. embarque 14. emboque 15. encasque 16. enfoque 17. ensaque 18. entrechoque 19. pique 20. quique 21. reboque 22. rebusque 23. recalque 24. repique 25. requeime 26. retoque 27. retruque 28. saque 29. soque 30. tique 31. toque 32. tremelique 33. tutuque	1. alcance 2. alce 3. avance 4. desenlace 5. disfarce 6. enlace 7. entrelace 8. lance 9. relance 10. retorce	1. alinde 2. desbunde 3. deslinda 4. engorde 5. envide 6. retarde 7. revide	1. desaguaxe 2. desencaixe 3. encaixe 4. enfaixe 5. rebaixe 6. relaxe

39. porte 40. reajuste 41. rebate 42. recorte 43. reimplante 44. remate 45. remonte 46. reparte 47. repinte 48. requinte 49. resgate 50. ressalte 51. tilinte 52. transplante 53. transporte				
--	--	--	--	--

SSE	ME	BE	GUE	LHE
1. emposse 2. endosse 3. passe 4. repasse 5. traspasse 6. trespasse	1. derrame 2. desarme 3. desmame 4. esparrame 5. informe 6. queime	1. arribe 2. bebe 3. derrube 4. desabe	1. derengue 2. desengasgue 3. engasgue 4. enxágue	1. abate 2. desencalhe 3. encalhe

TRE	BRE	NE	LE	JE	SE
1. alvitre 2. deslastre 3. deslustre	1. desdobre 2. dobre	1. encarne	1. desfile	1. entraxe	1. envase

APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA O EXPERIMENTO DE PSEUDOPALAVRAS

Termo de Consentimento Livre e Esclarecido para o experimento de pseudopalavras

Programa de Pós-graduação em Letras UFRGS
Estudos da Linguagem: Fonologia e Morfologia

O Sr.(a) está sendo convidado(a) a participar do projeto de pesquisa intitulado **Análise da produtividade do processo morfológico de derivação regressiva por meio de experimento com pseudopalavras**, sob a responsabilidade do pesquisador e orientador Dr. Luiz Carlos da Silva Schwindt e condução do pesquisador Eduardo Elisalde Toledo, doutorando do curso de pós-graduação em Letras. Para participar, você precisa ter 18 anos ou mais e estar cursando graduação em alguma universidade ou faculdade brasileira.

O objetivo deste trabalho é avaliar os fatores linguísticos envolvendo a produção de formas derivadas no português brasileiro. Para realizar o estudo, será necessário que o(a) Sr.(a) se disponibilize a participar no fornecimento de respostas acerca de qual forma derivada ou de qual forma infinitiva é a mais adequada no contexto frasal fornecido pelo experimento, ambas as tarefas envolvem palavras inventadas. Para a instituição e para sociedade, esta pesquisa fornecerá dados que serão comparados com outras fontes como dicionários e *corpus* linguístico.

A pesquisa é feita de forma voluntária e não apresenta remuneração ou algum benefício direto, material ou financeiro. Não haverá custos financeiros para os participantes desta pesquisa. Quanto aos riscos, a participação na pesquisa não implicará nenhum dano moral aos participantes; porém, poderá haver riscos mínimos relacionados a cansaço, tédio e constrangimento. O tempo previsto para conclusão do experimento é de dez minutos.

Em virtude das informações coletadas serem utilizadas unicamente com fins científicos, garantimos o total sigilo e confidencialidade, através da assinatura deste termo, a que o(a) Sr.(a) tem acesso inclusive para download. Além disso, o(a) Sr.(a) terá o 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. Caso você queira conhecer os resultados desta pesquisa, estes ficarão à disposição com os responsáveis e, posteriormente, serão publicados como parte integrante da tese de doutorado.

Para qualquer esclarecimento no decorrer da sua participação, favor contatar o pesquisador assistente (eduardoelisalde@gmail.com) ou o pesquisador responsável (schwindt@ufrgs.br) 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!

APÊNDICE C – EXPERIMENTO COM PSEUDOPALAVRAS

Experimento com pseudopalavras

*Obrigatório

1. Olá! Gostaríamos que você participasse de uma pesquisa que estamos realizando sobre processos morfológicos de formação de palavras no português brasileiro, que é parte de um projeto de pesquisa de doutorado da área de Letras, na UFRGS. São seis questões de múltipla escolha em que não há uma resposta correta, porque envolvem palavras inventadas. Então não se preocupe em "acertar"; queremos que você escolha a forma que "soe melhor", que pareça mais natural para você. Garantimos que levará, no máximo, cinco minutos para responder as perguntas que apresentaremos. Clique no link abaixo para ler em detalhes o Termo de Consentimento Livre e Esclarecido para participar do experimento.

https://docs.google.com/document/d/18JLLaH62cV8D4HSsZv7_GgH5hmLYY8BTj9YjVp9imj8/edit?usp=sharing *

Marcar apenas uma oval.

Eu desejo participar do experimento.

Eu não desejo participar do experimento. *Pare de preencher este formulário.*

Informação prévia

Antes de começar o nosso experimento, responda algumas perguntas.

2. **Curso de graduação:** *

3. **Sexo:** *

Marcar apenas uma oval.

Feminino

Masculino

Outro:

4. **Idade:** *

Questão 01

5. João costuma MALUTER diariamente. Ele, contudo, precisa comer muito antes de

.....

*

Marcar apenas uma oval.

sua malute

seu malute

sua maluta

seu maluto

Questão 02

16/01/2020 Experimento com pseudopalavras

https://docs.google.com/forms/d/1j8JK3_PXKG1oOy3W21YIdZadxbwe2AqEf-nyNSAq3Gg/edit 2/3

6. **Pedro tenta FETADIR sempre antes de dormir. Seu amigo recomendou para melhorar sua concentração. ***

Marcar apenas uma oval.

- essa fetade
- essa fetada
- esse fetado
- esse fetade

Questão 03

7. **Maria não tem medo de MOPITAR enquanto dirige. Mas já foi avisada que pode ser um crime grave de trânsito. ***

Marcar apenas uma oval.

- o mopito
- a mopita
- o mopite
- a mopite

Questão 04

8. **Nos aposentos da casa, devido ao excesso de umidade, somente com ZENETO mantemos a casa limpa. Se esquecemos de , sofreremos com alergias o ano inteiro. ***

Marcar apenas uma oval.

- zeneter
- zenetir
- zenetar

Questão 05

9. **Cada vez mais comum nas democracias do mundo, optar pela BOPLIDA é uma prática bastante popular entre o público jovem. Se não puderem se sentem antiquados. ***

Marcar apenas uma oval.

- boplider
- boplidir
- boplidir

Questão 06

10. **Para garantir a privacidade das mensagens eletrônicas, executa-se o FITUQUE antes de seu envio. Ao logar no aplicativo, é só uma vez para garantir que as mensagens sejam protegidas de invasores. ***

Marcar apenas uma oval.

- fituquir
- fitucar
- fituquer

APÊNDICE D – QUADRO DE CÓDIGOS DOS GRUPOS DE FATORES LINGUÍSTICOS

(1) VOGAL TEMÁTICA	CÓDIGO
<i>verbos da 1ª conjugação</i>	a
<i>verbos da 2ª conjugação</i>	e
<i>verbos da 3ª conjugação</i>	i
(2) MANUTENÇÃO DA VOGAL TEMÁTICA	CÓDIGO
<i>preserva a vogal temática da base</i>	s
<i>não preserva a vogal temática da base</i>	
(5) COMPLEXIDADE MORFOLÓGICA	CÓDIGO
<i>palavra simples</i>	l
<i>palavra prefixada</i>	m
<i>palavra sufixada</i>	n
(4) VOGAL MÉDIA TÔNICA	CÓDIGO
<i>presença da vogal média tônica</i>	j
<i>ausência da vogal média tônica</i>	k
(5) GÊNERO GRAMATICAL	
<i>gênero masculino</i>	m
<i>gênero feminino</i>	f
(6) NÚMERO DE SÍLABAS	CÓDIGO
<i>monossílabo</i>	1
<i>dissílabo</i>	2
<i>trissílabo</i>	3
<i>polissílabo</i>	4
(6) ONSET DA ÚLTIMA SÍLABA	CÓDIGO
<i>presença de onset</i>	5
<i>ausência de onset</i>	6

(8) CONTEXTO FONOLÓGICO PRECEDENTE (modo)	CÓDIGO
<i>nasal</i>	r
<i>plosiva</i>	b
<i>fricativa</i>	t
<i>africada</i>	u
<i>lateral</i>	v
<i>tepe</i>	x
<i>vogal</i>	z
(9) CONTEXTO FONOLÓGICO PRECEDENTE (ponto)	CÓDIGO
<i>bilabial</i>	X
<i>labio-dental</i>	V
<i>alveolar</i>	U
<i>palatal</i>	T
<i>velar</i>	S
<i>vogal alta</i>	#
<i>vogal média</i>	\$
<i>vogal baixa</i>	%

APÊNDICE E – QUADRO DE RESPOSTAS DO EXPERIMENTO

Masculin o	18	seu malute	essa fetade	a mopita	zenetar	boplidar	fitucar
Masculin o	19	sua maluta	essa fetada	o mopite	zeneter	boplidar	fitucar
Feminin o	19	sua maluta	essa fetada	o mopito	zeneter	boplidar	fituquir
Feminin o	19	sua maluta	essa fetada	o mopito	zeneter	boplidar	fitucar
Feminin o	20	sua maluta	esse fetade	o mopito	zenetar	boplidar	fitucar
Feminin o	20	seu malute	esse fetado	o mopite	zenetar	boplidir	fitucar
Feminin o	20	sua maluta	essa fetade	o mopito	zenetar	boplidir	fitucar
Feminin o	20	sua malute	essa fetada	o mopito	zenetar	boplidir	fitucar
Feminin o	21	sua malute	essa fetada	o mopite	zenetir	boplider	fitucar
Feminin o	21	sua maluta	essa fetada	o mopito	zeneter	boplidar	fitucar
Feminin o	21	seu maluto	esse fetado	o mopito	zenetar	boplidar	fitucar
Masculin o	21	sua maluta	essa fetade	a mopite	zenetar	boplidar	fitucar
Feminin o	21	seu malute	esse fetado	o mopito	zenetar	boplidar	fitucar
Feminin o	21	sua malute	essa fetada	o mopito	zenetar	boplidar	fituquir
Feminin o	21	seu malute	essa fetada	o mopito	zenetar	boplidar	fitucar

Feminin o	21	seu malute	essa fetade	o mopite	zenetar	boplidir	fitucar
Feminin o	21	seu malute	esse fetado	o mopito	zenetar	boplidar	fitucar
Masculin o	21	sua maluta	esse fetado	o mopite	zenetar	boplidar	fitucar
Feminin o	22	seu malute	essa fetada	o mopite	zenetar	boplidir	fitucar
Feminin o	22	seu maluto	esse fetado	o mopito	zenetar	boplidar	fitucar
Masculin o	22	seu malute	essa fetade	o mopite	zenetir	boplidir	fitucar
Feminin o	22	sua maluta	esse fetado	o mopite	zenetar	boplidir	fitucar
Feminin o	22	seu malute	esse fetade	o mopite	zenetar	boplidar	fitucar
Masculin o	22	seu malute	esse fetado	a mopita	zenetar	boplidir	fitucar
Feminin o	23	seu malute	esse fetade	o mopito	zenetar	boplidar	fituquir
Feminin o	23	seu malute	essa fetada	o mopito	zenetir	boplidar	fituquir
Masculin o	23	seu maluto	essa fetada	o mopite	zenetar	boplidar	fitucar
Feminin o	24	seu malute	essa fetade	o mopite	zenetar	boplidar	fituquir
Masculin o	25	seu malute	esse fetade	o mopite	zenetar	boplidar	fitucar
Feminin o	25	sua maluta	esse fetade	o mopite	zenetar	boplidir	fitucar
Feminin o	25	sua malute	essa fetada	a mopita	zenetar	boplidar	fituquir

Feminin o	26	seu malute	esse fetade	o mopito	zeneter	boplider	fitucar
Feminin o	26	seu malute	esse fetado	a mopita	zenetar	boplidar	fitucar
Feminin o	26	seu malute	essa fetada	a mopita	zenetar	boplidir	fitucar
Feminin o	26	sua maluta	esse fetade	o mopito	zenetar	boplidar	fitucar
Masculin o	26	sua maluta	essa fetade	a mopita	zenetar	boplidar	fitucar
Feminin o	26	seu malute	esse fetade	a mopite	zeneter	boplidir	fituquer
Masculin o	27	sua maluta	essa fetade	a mopite	zenetar	boplidar	fitucar
Masculin o	28	sua maluta	esse fetade	a mopita	zenetar	boplidar	fitucar
Feminin o	28	seu malute	essa fetade	o mopite	zenetar	boplidar	fitucar
Feminin o	28	sua maluta	essa fetada	o mopite	zeneter	boplidar	fitucar
Masculin o	28	sua maluta	esse fetado	o mopito	zenetar	boplidar	fitucar
Feminin o	28	sua malute	essa fetada	o mopito	zenetar	boplidar	fituquir
Feminin o	29	sua maluta	essa fetade	a mopita	zenetar	boplidar	fitucar
Feminin o	29	sua maluta	essa fetada	a mopita	zeneter	boplidar	fitucar
Masculin o	29	seu malute	essa fetada	o mopito	zenetar	boplidir	fituquir
Feminin o	29	seu malute	esse fetade	o mopito	zenetar	boplidar	fitucar

Feminin o	29	sua maluta	essa fetada	a mopita	zenetar	boplidir	fitucar
Masculin o	30	sua malute	essa fetade	a mopita	zenetar	boplidar	fitucar
Feminin o	30	seu malute	essa fetada	o mopito	zenetar	boplidar	fitucar
Feminin o	30	sua maluta	esse fetado	o mopite	zenetir	boplidar	fitucar
Feminin o	30	seu malute	essa fetada	o mopito	zenetar	boplidar	fitucar
Masculin o	30	seu malute	essa fetada	o mopito	zenetar	boplidir	fitucar
Masculin o	30	sua maluta	essa fetada	o mopito	zenetar	boplider	fitucar
Feminin o	32	sua maluta	essa fetade	o mopito	zenetir	boplidar	fitucar
Feminin o	33	sua malute	essa fetada	o mopite	zenetar	boplidir	fitucar
Feminin o	33	seu malute	essa fetade	o mopite	zenetar	boplidar	fitucar
Feminin o	33	seu maluto	essa fetada	a mopita	zenetar	boplidar	fitucar
Feminin o	34	sua malute	essa fetada	o mopite	zenetir	boplidar	fitucar
Feminin o	34	sua maluta	essa fetada	o mopito	zenetar	boplider	fitucar
Feminin o	34	seu malute	esse fetado	o mopito	zenetar	boplidar	fitucar
Masculin o	35	sua maluta	essa fetada	o mopite	zenetar	boplidar	fitucar
Feminin o	35	sua maluta	essa fetada	o mopito	zenetar	boplidar	fitucar

Feminin o	35	seu malute	essa fetada	a mopita	zenetar	boplidar	fitucar
Feminin o	35	seu malute	esse fetade	o mopito	zenetar	boplidar	fitucar
Feminin o	35	seu malute	esse fetado	a mopite	zenetar	boplidar	fituquir
Feminin o	37	sua maluta	essa fetade	o mopite	zenetar	boplidir	fitucar
Feminin o	37	seu malute	essa fetada	o mopito	zenetar	boplidar	fitucar
Feminin o	37	sua malute	essa fetade	o mopite	zenetar	boplidar	fitucar
Feminin o	37	sua maluta	essa fetada	a mopite	zeneter	boplider	fitucar
Masculin o	37	sua malute	esse fetade	o mopito	zenetar	boplidar	fituquer
Feminin o	38	sua malute	essa fetada	a mopite	zenetar	boplidar	fitucar
Feminin o	38	sua maluta	essa fetada	o mopito	zenetar	boplidar	fitucar
Feminin o	39	seu maluto	essa fetade	a mopite	zenetar	boplidar	fitucar
Masculin o	39	sua maluta	essa fetada	o mopite	zenetar	boplidar	fitucar
Feminin o	40	sua maluta	essa fetade	o mopito	zenetar	boplidar	fituquir
Feminin o	40	sua maluta	essa fetade	o mopito	zenetar	boplidir	fituquir
Feminin o	42	seu malute	esse fetado	o mopito	zenetar	boplider	fituquir
Feminin o	42	seu malute	essa fetade	a mopita	zenetar	boplidar	fitucar

Masculin o	42	seu malute	esse fetade	a mopita	zenetar	boplidar	fitucar
Feminin o	43	sua maluta	esse fetado	o mopite	zenetar	boplidir	fituquer
Feminin o	43	seu malute	essa fetada	o mopite	zenetar	boplidar	fitucar
Feminin o	44	seu malute	esse fetade	o mopite	zenetar	boplidar	fituquir
Feminin o	45	sua maluta	esse fetade	a mopite	zenetar	boplidar	fitucar
Feminin o	46	seu malute	essa fetada	o mopite	zenetar	boplidir	fitucar
Masculin o	46	seu maluto	essa fetade	a mopite	zenetar	boplidar	fitucar
Feminin o	47	seu maluto	essa fetade	o mopito	zenetar	boplidar	fitucar
Feminin o	47	seu malute	essa fetada	o mopite	zenetar	boplidar	fitucar
Feminin o	48	seu malute	esse fetado	o mopite	zenetar	boplidir	fitucar
Feminin o	48	sua maluta	esse fetado	o mopite	zenetar	boplidir	fitucar
Feminin o	50	sua maluta	esse fetade	o mopito	zenetar	boplidar	fitucar
Masculin o	51	sua maluta	essa fetade	a mopita	zeneter	boplidar	fituquir
Feminin o	54	sua maluta	essa fetade	o mopito	zenetar	boplidar	fitucar
Masculin o	54	sua malute	essa fetade	o mopito	zenetar	boplidar	fitucar
Feminin o	55	sua malute	essa fetada	a mopite	zenetar	boplidar	fitucar

Feminin o	56	sua maluta	essa fetada	o mopito	zenetar	boplidar	fituquir
Masculin o	56	sua maluta	essa fetada	a mopita	zenetar	boplidar	fitucar
Feminin o	57	sua maluta	essa fetade	o mopito	zenetar	boplidir	fituquir
Feminin o	57	seu malute	essa fetade	a mopita	zenetir	boplidar	fitucar
Masculin o	61	sua maluta	essa fetade	o mopito	zenetar	boplidar	fitucar

APÊNDICE F – QUADRO DE DISTRIBUIÇÃO DAS RESPOSTAS DOS INFORMANTES AO EXPERIMENTO ONLINE SEGUNDO OS FATORES SOCIAIS

Masculino				Feminino			
Questão 01	18-30 anos	31-50 anos	51-61 anos	Questão 01	18-30 anos	31-50 anos	51-61 anos
a	50% (8)	40% (2)	75% (3)	a	34% (13)	37% (12)	60% (3)
e(M)	38% (6)	20% (1)	0% (0)	e(M)	47% (18)	41% (13)	20% (1)
e(F)	6% (1)	20% (1)	25% (1)	e(F)	13% (5)	13% (4)	20% (1)
o	6% (1)	20% (1)	0% (0)	o	6% (2)	9% (3)	0% (0)
Questão 02	18-30 anos	31-50 anos	51-61 anos	Questão 02	18-30 anos	31-50 anos	51-61 anos
a	31% (5)	40% (2)	25% (1)	a	45% (17)	41% (13)	40% (2)
e(M)	12% (2)	40% (2)	0% (0)	e(M)	21% (8)	13% (4)	0% (0)
e(F)	38% (6)	20% (1)	75% (3)	e(F)	13% (5)	28% (9)	60% (3)
o	19% (3)	0% (0)	0% (0)	o	21% (8)	18% (6)	0% (0)
Questão 03	18-30 anos	31-50 anos	51-61 anos	Questão 03	18-30 anos	31-50 anos	51-61 anos
a	31% (5)	20% (1)	50% (2)	a	16% (6)	9% (3)	20% (1)
e(M)	31% (5)	40% (2)	0% (0)	e(M)	29% (11)	15% (5)	0% (0)
e(F)	12% (2)	20% (1)	0% (0)	e(F)	3% (1)	38% (12)	20% (1)
o	26% (4)	20% (1)	50% (2)	o	52% (20)	38% (12)	60% (3)
Questão 04	18-30 anos	31-50 anos	51-61 anos	Questão 04	18-30 anos	31-50 anos	51-61 anos

ar	88% (14)	100% (5)	75% (3)	ar	74% (28)	91% (29)	80% (4)
er	6% (1)	0% (0)	25% (1)	er	18% (7)	3% (1)	0% (0)
ir	6% (1)	0% (0)	0% (0)	ir	8% (3)	6% (2)	20% (1)
Questão 05	18-30 anos	31-50 anos	51-61 anos	Questão 05	18-30 anos	31-50 anos	51-61 anos
ar	68% (11)	100% (5)	100% (4)	ar	68% (26)	69% (22)	80% (4)
er	6% (1)	0% (0)	0% (0)	er	6% (2)	9% (3)	0% (0)
ir	26% (4)	0% (0)	0% (0)	ir	26% (10)	22% (7)	20% (1)
Questão 06	18-30 anos	31-50 anos	51-61 anos	Questão 06	18-30 anos	31-50 anos	51-61 anos
ar	88% (14)	80% (4)	75% (3)	ar	79% (30)	81% (26)	60% (3)
er	6% (1)	20% (1)	0% (0)	er	3% (1)	3% (1)	0% (0)
ir	6% (1)	0% (0)	25% (1)	ir	18% (7)	16% (5)	40% (2)
	Masculino	Feminino		Faixa Etária			
Questão 01				Questão 01	18-30 anos	31-50 anos	51-61 anos
a	52% (13)	37% (28)		a	39% (21)	38% (14)	67% (6)
e(M)	28% (7)	43% (32)		e(M)	44% (24)	38% (14)	11% (1)
e(F)	12% (3)	13% (10)		e(F)	11% (6)	13% (5)	22% (2)
o	8% (2)	7% (5)		o	6% (3)	11% (4)	0% (0)
Questão 02				Questão 02	18-30 anos	31-50 anos	51-61 anos
a	32% (8)	43% (32)		a	41% (22)	40% (15)	33% (3)

e(M)	16% (4)	16% (12)		e(M)	18% (10)	16% (6)	0% (0)
e(F)	40% (10)	23% (17)		e(F)	20% (11)	27% (10)	67% (6)
o	12% (3)	18% (14)		o	20% (11)	16% (6)	0% (0)
Questão 03				Questão 03	18-30 anos	31-50 anos	51-61 anos
a	32% (8)	13% (10)		a	20% (11)	11% (4)	33% (3)
e(M)	28% (7)	21% (16)		e(M)	24% (16)	67% (25)	0% (0)
e(F)	12% (3)	19% (14)		e(F)	6% (3)	35 % (13)	12% (1)
o	28% (7)	47% (35)		o	44% (24)	35% (13)	55% (5)
Questão 04				Questão 04	18-30 anos	31-50 anos	51-61 anos
ar	88% (22)	81% (61)		ar	78% (42)	92% (34)	76% (7)
er	8% (2)	11% (8)		er	15% (8)	3% (1)	12% (1)
ir	4% (1)	8% (6)		ir	7% (4)	5% (2)	12% (1)
Questão 05				Questão 05	18-30 anos	31-50 anos	51-61 anos
ar	80% (20)	69% (52)		ar	68% (37)	73% (27)	88% (8)
er	4% (1)	7% (5)		er	6% (3)	8% (3)	0% (0)
ir	16% (4)	24% (18)		ir	26% (14)	19% (7)	12% (1)
Questão 06				Questão 06	18-30 anos	31-50 anos	51-61 anos
ar	84% (21)	78% (59)		ar	81% (44)	81% (30)	67% (6)

er	8% (2)	2% (2)		er	4% (2)	5% (2)	0% (0)
ir	8% (2)	20% (14)		ir	15% (8)	14% (5)	33% (3)

APÊNDICE G – LISTA DE FREQUÊNCIA DOS NOMES DEVERBAIS

trabalho	815927
acordo	337008
ensino	320178
apoio	161896
desempenho	133520
consumo	126284
encontro	66479
erro	51916
experimento	50524
retorno	44273
desafio	40471
manejo	36828
cultivo	34806
comando	33133
avanço	31995
atraso	26478
preparo	25043
apelo	24879
confronto	22755
protesto	20538
abandono	17905
treino	17758
bloqueio	16358
trato	14585
tiro	13145
conforto	12937
repouso	12250
desconto	11697
assalto	10993
roubo	10224

envio	9966
pastejo	9737
transtorno	9296
passeio	9131
custeio	8980
socorro	8483
acerto	8412
desespero	7424
enxerto	6841
descanso	6512
manuseio	6312
sorteio	6190
recuo	5806
sustento	5726
castigo	5284
aviso	5255
reparo	5206
grito	4788
exagero	4774
amparo	4655
enterro	4592
retardo	4280
despacho	4179
choro	4112
cerco	3643
namoro	3594
espanto	3534
mergulho	3360
pouso	3286
apego	3252
sopro	3022
anseio	2559
despejo	2524

suborno	2524
disparo	2458
aperto	2418
encanto	2199
bombardeio	2177
agravo	2168
conserto	1915
confisco	1876
sossego	1818
consolo	1796
descuido	1640
refino	1635
estrago	1481
rateio	1371
nado	1367
desembolso	1356
abalo	1327
desabafo	1310
arremesso	1226
expurgo	1057
desencontro	1024
desencanto	988
pulo	952
veraneio	932
sobressalto	862
cotejo	767
arrasto	754
arrepio	745
desleixo	730
desagrado	725
tombo	722
fabrico	718
resguardo	703

endosso	695
assombro	652
refugo	650
embalo	644
aceno	614
ressalto	587
extravio	583
arremedo	529
apuro	527
festejo	523
restauro	521
desacato	509
tropeço	496
assomo	482
lanço	454
arrojo	452
empuxo	447
malogro	443
transbordo	416
desbloqueio	410
agito	386
tino	381
desassossego	364
desmaio	346
desterro	321
atropelo	317
desmando	305
arresto	299
descortino	292
descaminho	284
esbulho	265
reclamo	260
degelo	257

espirro	242
bochecho	241
cochilo	223
arroubo	218
desafogo	215
afago	204
berro	195
assobio	178
rutilo	178
olvido	168
bocejo	160
talho	154
conchavo	150
escracho	147
respiro	147
voo	147
gracejo	137
repto	134
gargarejo	132
reconto	132
arrumo	131
desengano	126
reenvio	106
abordo	98
aprumo	92
calço	92
cochicho	87
rogo	80
urro	75
carreto	74
repuxo	74
arranco	74
arroto	71

confesso	67
silvo	65
aboio	63
acalanto	60
quebranto	60
sorvo	59
desdouro	57
mexerico	55
rodopio	55
galanteio	54
respingo	53
converso	51
rojo	47
volteio	47
menoscabo	42
resmungo	41
xingo	40
meneio	39
acrescento	38
rebaixo	36
carrego	35
enleio	34
esculacho	32
desabono	31
remelexo	31
floreio	29
gorjeio	28
desaperto	26
requebro	23
almejo	21
retraso	21
ceceio	20
desbarato	20

desembargo	20
desmancho	20
descaro	19
aferro	18
esconjuro	18
bufo	17
esbarro	17
maneio	17
esparramo	16
arreganho	15
embolso	15
desfruto	15
desaponto	14
desentulho	14
conjuro	13
regateio	13
apronto	12
ameaço	12
recolho	12
apresto	12
avio	11
desenredo	11
remolho	11
amuo	9
ofego	9
calafeto	8
desbordo	8
acúmulo	7
arrufo	7
cacarejo	7
conchego	7
remedo	7
arrulho	6

gaguejo	6
ralho	6
abafo	5
acanho	5
afogo	5
alinho	5
desaterro	5
granjeio	5
partejo	5
rebusno	5
pipoco	5
saracoteio	5
soçobro	5
adejo	4
antecipo	4
latejo	4
saqueio	4
transvio	4
trino	4
deslastro	4
restilo	4
arquejo	3
câmbio	3
enxovalho	3
gorgolejo	3
pipio	3
recobro	3
toureio	3
tresmalho	3
trincho	3
costeio	3
desgarro	3
traspasso	3

alívio	2
coaxo	2
desensino	2
desgabo	2
pestanejo	2
rastejo	2
regelo	2
vozeio	2
zumbo	2
saçarico	2
encalho	2
boquejo	1
cantarejo	1
chilreio	1
desapoio	1
desconchavo	1
desdoiro	1
garganteio	1
mio	1
pipilo	1
refreio	1
relampejo	1
sapateio	1
trovisco	1
ululo	1
zangarreio	1
aforro	1
arregaço	1
arrenego	1
derriço	1
pareio	1

reforma	111907
amostra	95298
perda	88752
escolha	84982
volta	82414
demanda	64911
compra	55948
troca	55948
ajuda	53523
disputa	41570
visita	36677
emenda	36427
procura	29567
fala	27954
consulta	26701
vaga	23951
entrega	23049
conversa	22418
espera	21329
quebra	18805
fuga	17658
mostra	16886
marcha	12984
pesca	12054
queima	8662
queixa	8624
recusa	8209
escuta	5255
recompensa	4163
remessa	3891

revenda	3487
barganha	3333
poda	3260
lavra	2860
mira	2741
desmama	2688
ordenha	2670
sobra	2550
resenha	2432
outorga	1885
risca	1588
afronta	1547
engorda	1370
permuta	1194
desova	1185
penhora	1059
recria	1035
pega	887
reconquista	817
paga	773
chancela	768
desfolha	749
surra	725
contenda	719
lida	646
reza	574
trava	545
avança	532
cata	502
salga	494
rega	490
espreita	427
capina	393

recolha	381
limpa	367
empurra	346
recaptura	339
torra	335
derrama	315
degola	309
talha	287
desforra	205
despesca	201
conjura	172
escusa	167
debulha	158
desobriga	145
corta	136
cora	125
sova	123
cabruca	115
empreita	96
apanha	95
delonga	84
rebaixa	81
amontoa	72
pendura	62
tosquia	56
sega	47
tosa	43
tosa	43
foda	40
tolda	33
pisa	32
referenda	32
monda	32

desafronta	27
destroca	23
gira	21
denúncia	17
ferra	17
achega	16
peleja	14
ronca	13
veda	12
esfrega	10
encolha	9
renúncia	9
escalda	9
pernoita	8
arrelia	6
charla	6
parla	6
pronúncia	6
chega	4
joeira	4
respiga	4
sapeca	3
dependura	2
reestampa	2
repergunta	2
solta	2
vareja	2
rebusca	2
desolha	1
embroma	1
esgana	1
repisa	1

debate	70148
corte	65314
combate	46174
ataque	41061
ajuste	33123
reajuste	25109
alcance	24488
enfoque	22650
resgate	15933
transplante	13219
empate	12878
desfile	11801
desgaste	10526
toque	9782
aparte	9444
abate	8962
saque	5310
embarque	4251
escape	3299
derrame	3038
desembarque	2697
desmonte	2591
levante	1796
enlace	1720
desfalque	1279
recalque	1250
disfarce	1196
encaixe	1193
desajuste	1186
informe	1165
desempate	1125

pique	982
requinte	854
deslize	803
desenlace	762
repique	584
rebate	483
retoque	469
desfrute	406
relance	378
remate	370
enfarte	354
arranque	329
revide	247
pernoite	242
encalhe	241
deslinde	212
trespasse	203
tique	187
decalque	183
entrechoque	161
desbunde	142
relaxe	102
desquite	95
descasque	95
traspasse	68
engaste	65
retarde	59
derrube	56
soque	53
engorde	40
remonte	25
quique	20
ressalte	20

desengate	18
tremelique	9
desgarre	8
despique	6
debique	5
desbarate	4
esparrame	2
retruque	1
arribe	1
requeime	1

ANEXO A – ÚLTIMO PARECER DO COMITÊ DE ÉTICA EM PESQUISA

DADOS DO PARECER

Número do Parecer: 3.824.787

Apresentação da Notificação:

Trata-se de relatório final de pesquisa conduzida pelos pesquisadores Luiz Carlos da Silva Schwindt e Eduardo Elisalde Toledo no período de 01/03/2019 a 29/12/2019.

Objetivo da Notificação:

O objetivo do relatório é apresentar as atividades desenvolvidas pelos pesquisadores e os principais resultados da pesquisa.

Avaliação dos Riscos e Benefícios:

Não se aplica

Comentários e Considerações sobre a Notificação:

Conforme o previsto, o experimento com pseudopalavras foi realizado com a participação de 100

Continuação do Parecer: 3.824.787

(cem) sujeitos. A proposta inicial era usar o do site Survey Monkey; porém, a plataforma Google Forms foi utilizada, por apresentar uma série de vantagens para aplicação do questionário online no formato desejado, segundo o pesquisador principal.

Considerações sobre os Termos de apresentação obrigatória:

Não se aplica

Conclusões ou Pendências e Lista de Inadequações:

O relatório está redigido de forma clara e completa, apresentando uma modificação feita em relação ao projeto inicial, bem como as principais atividades desenvolvidas e os resultados alcançados.

Considerações Finais a critério do CEP:

APROVADO.