

Vogais intervenientes entre raiz e sufixos iniciados por vogal em português brasileiro

Intervening vowels between root and vowel-initial suffix in Brazilian Portuguese

Luiz Carlos Schwindt¹

Camila Witt Ulrich²

Nathan Santos Barcellos³

RESUMO: neste trabalho, descreve-se o comportamento de palavras derivadas pelos sufixos -al e -oso do português brasileiro com foco nos casos em que eventualmente emerge uma vogal entre raiz e sufixo (ex. rac+i+al, grac+i+oso). Com objetivo de contribuir para o debate sobre o domínio da sufixação na língua, analisam-se quantitativamente todos os adjetivos e substantivos listados no Dicionário Aurélio Eletrônico. Os resultados sugerem que o fenômeno, bastante escasso na língua, tem bases histórico-lexicais mais do que estruturais, figurando como exceção ao padrão geral observado nos dados de morfologia sufixal alinhada à raiz.

PALAVRAS-CHAVE: raiz; tema; palavra; derivação sufixal.

ABSTRACT: this article addresses the behavior of words derived by the suffixes -al and -oso in Brazilian Portuguese, focusing on cases in which a vowel can emerge between root and suffix (e.g., rac+i+al, grac+i+oso). In order to contribute to the debate on the domain of suffix derivation in the language, all adjectives and nouns listed in the Aurélio Eletrônico dictionary are analyzed from a quantitative perspective. The results suggest that the phenomenon, scarcely attested in the language, is historically and lexically rather than structurally based, and can be considered an exception to the general pattern of root-aligned morphology observed in the data.

KEYWORDS: root; stem; word; suffixation.

¹ Professor na Universidade Federal do Rio Grande do Sul, doutor pela Pontifícia Universidade Católica do Rio Grande do Sul, bolsista de produtividade do CNPq 309576/2022-9, schwindt@ufrgs.br, ORCID <https://orcid.org/0000-0003-0533-589X>.

² Professora na Universidade Federal do Pampa/campus Jaguarão, doutora pela Universidade Federal do Rio Grande do Sul, camilawittulrich@gmail.com, ORCID <https://orcid.org/0000-0003-1161-0472>.

³ Graduado em Letras na Universidade Federal do Rio Grande do Sul, nathan.s.barcellos@gmail.com, ORCID <https://orcid.org/0000-0001-9703-3850>.

1 Introdução

Neste artigo apresentamos resultados de uma pesquisa quantitativa sobre palavras dicionarizadas do português brasileiro (PB) em que uma vogal — aqui denominada vogal interveniente (VI) — se apresenta entre a raiz e os sufixos iniciados por vogal -al e -oso. A questão não é trivial, uma vez que há casos, exemplificados a seguir, em que nesse contexto morfológico: nenhuma vogal emerge, (1); vogais foneticamente iguais ou muito semelhantes à última vogal da palavra emergem, (2); vogais foneticamente diferentes da última vogal da palavra emergem, (3).

(1)	a.	canal	*canoal	can+o+al
	b.	chuvoso	*chuvaoso	chuv+a+oso
(2)	a.	facial	*façal	fac+i+al
	b.	afetuoso	*afetoso	afet+u+oso
(3)	a.	bestial	*bestal	best+i+al
	b.	precioso	*preçoso	prec+i+oso

A pergunta de base desta investigação diz respeito ao domínio da derivação sufixal, se a raiz ou o tema/palavra (raiz+vogal temática). Harris (1983), em relação ao espanhol, assume a primeira perspectiva, que é contraditada por Bermúdez-Otero (2006, 2013), por exemplo. Já para o PB, a primeira perspectiva é a assumida por Schwindt (2013) e Ulrich & Schwindt (2018), entre outros, em contraste ao que sustentam Câmara Jr. (1977) e Matzenauer & Bisol (2016), entre outros.

Ainda que esta análise contribua diretamente para esse debate sobre o domínio da derivação, o objetivo deste texto é principalmente descritivo. Nesse sentido, partimos de um levantamento de dados em que uma vogal se situa entre raiz e sufixo em adjetivos e substantivos do PB e quantificamos possíveis

condicionadores para a emergência desse segmento. Nesse sentido, as variáveis *segmento final da base*, *ponto de articulação da última consoante da base*, *presença de consoante epentética*, *número de sílabas da base*, *sufixo*, *alomorfia da base*, *frequência lexical da palavra derivada* e *datação* são examinadas do ponto de vista estatístico-descritivo.

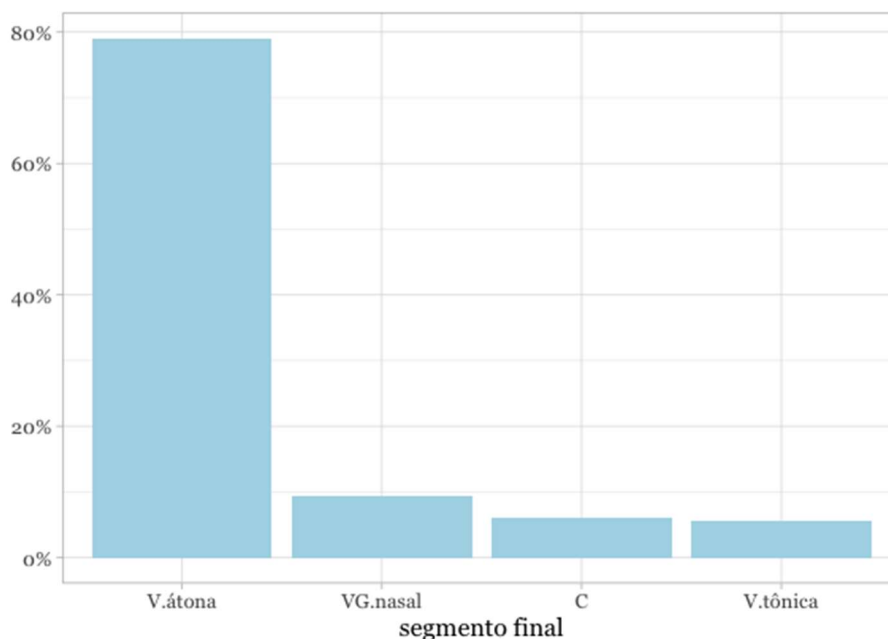
O artigo se organiza como segue. Na seção 2, sintetizamos as principais ideias sobre a controvérsia morfologia baseada na raiz *versus* morfologia baseada no tema/palavra. Na seção 3, descrevemos brevemente a metodologia de nossa análise. Na seção 4, apresentamos os principais resultados acompanhados de discussão. Na seção 5, por fim, estão nossas considerações finais.

2 Morfologia baseada na palavra *versus* morfologia baseada na raiz

Embora conte com palavras fechadas por consoante, de modo especial em nível de superfície, o português, como outras línguas românicas, apresenta, segundo o Dicionário Aurélio Eletrônico versão 2010, clara preferência por palavras fechadas por vogais átonas; em seguida, com índices bastante inferiores, a ordem de ocorrência identificada é, respectivamente, de ditongos nasais, consoantes e vogais tônicas.⁴

⁴ Neste gráfico estão amalgamadas por conveniência, na categoria V.átona, substantivos terminados em vogais átonas (78,5%) e em ditongos orais (0,5%), e, na categoria VG.nasal, substantivos terminados em ditongos nasais (9%) e em vogais nasais (0,4%).

Gráfico 1 – Segmento em final de substantivos em PB – Dicionário Aurélio



Fonte: Adaptado de Schwindt (2020, p. 286).

Se considerarmos a análise de Câmara Jr. (1970), as três primeiras colunas do gráfico acima dizem respeito a vocábulos temáticos, enquanto a última corresponde aos atemáticos, já que a vogal tônica não se junta à raiz como índice temático, mas é parte indissociável dela.

É a evidência de uma considerável maioria de palavras temáticas na língua que justifica a pergunta: afixos se ligam, então, diretamente à raiz – elemento despido de vogal temática (VT) – ou ao tema/palavra? A resposta a essa questão não é trivial e divide os linguistas. Aqui retomamos abordagens contraditórias envolvendo, principalmente, o espanhol e o português, dada a proximidade entre esses sistemas.

2.1 Morfologia baseada no tema/palavra

A postulação do tema/palavra como base mínima para derivações em línguas românicas foi defendida por diversos autores (CAMARA JR., 1977; SCALISE, 1984; BERMÚDEZ-OTERO, 2006, 2013; MATZENAUER & BISOL,

2016; entre outros). Destacamos aqui os principais argumentos de Bermúdez-Otero (2006), para o espanhol, e de Matzenauer & Bisol (2016), para o português.

Bermúdez-Otero (2006, 2013) problematiza o status das VTs em nomes e adjetivos do espanhol, uma vez que a distribuição dessas vogais não pode ser prevista nessa língua pela forma fonológica das raízes nem por traços gramaticais (por exemplo, os de gênero).⁵ Sugere, então, que, a rigor, o léxico armazena temas – raízes associadas a vogais temáticas. Como nem sempre a VT se exponencia, o autor lança mão de um processo fonológico geral e regular de apagamento de vogais finais inacentuadas antes de sufixos iniciados por vogal (ex. [mán-o] - [man-áθ-a]).

Algumas das vantagens da hipótese de armazenamento de temas elencadas pelo autor são: (i) a distribuição unificada de vogais temáticas de verbos, nomes e adjetivos; (ii) a explicação da alternância [o]/[we] no interior das raízes motivada pelo acento lexical (ex. contar ~ cuento). Além disso, ele argumenta que dados psicolinguísticos suportam a hipótese de armazenamento de duas formas alomórficas (ex. [kont-a] e [kwent-a], que cobrem tanto *contista* quanto *cuento*).

Matzenauer & Bisol (2016, p. 341) assumem ponto de vista análogo para o PB, isto é, defendem que “o tema está armazenado no léxico profundo”, numa gramática baseada em níveis, e que serve de base para novas derivações na língua. As autoras se valem de argumentos morfofonológicos, morfológicos e semânticos para sustentar esse ponto de vista. As evidências morfofonológicas residem nos processos de metafoia nominal (MIRANDA, 2002), que tem lugar a partir da anexação da vogal temática -o (ex. p[ɔ]rc- / p[o]rco), e de neutralização e elevação das vogais médias pretônicas (ex. b[ɔ]la / b[o]llada), dependentes esses últimos processos do preenchimento da VT para atribuição do acento primário.⁶ A evidência morfológica provém de formas derivadas por sufixo. Matzenauer & Bisol citam, entre outros, exemplos de palavras formadas pelos sufixos -inho e -ico, que têm força de preservar informações herdadas da vogal final do tema a

⁵ Bermúdez-Otero discute aspectos da seleção de bases alomórficas no espanhol a partir da Morfologia Distribuída (HALLE; MARANTZ, 1993) e da Teoria da Otimidade Estratal (KIPARSKY, 2000), assumindo a superioridade do último modelo sobre o primeiro.

⁶ Na visão das autoras, o acento primário é atribuído à palavra (BISOL, 1992). Há, também para o PB, análises que assumem a atribuição do acento no nível do radical (LEE, 1997).

que se ligam (como no caso do gênero em *bolo* / *bolinho* ou *verão* / *veranico*). À semelhança do que sustenta Bermúdez-Otero para o espanhol, neste caso se pressupõe, como sugeriu Câmara Jr., que um processo fonológico elide a VT antes da entrada desses sufixos. Além disso, sustentam que a VT é preservada no interior de alguns vocábulos derivados por -oso (ex. monstr-u-oso), -ão (ex. feij-o-ada) e sufixos nominalizadores (ex. alinh-a-mento) — recorte em que parcialmente se localiza a análise empreendida. A evidência semântica, por fim, diz respeito a oposições de significado entre palavras de raiz similar, como *fruto* e *fruta* ou *barco* e *barca*, identificadas apenas a partir da vogal final. Cada uma dessas palavras, por consequência, licencia o uso de diferentes sufixos para novas formações (ex. fruto / frutuoso; fruta / frutaria)⁷. Assim, a vogal final seria um requisito necessário para que as restrições de seleção dos sufixos sejam obedecidas.

2.2 Morfologia baseada na raiz

A postulação da raiz como base para derivações em línguas românicas foi defendida por diversos autores (HARRIS, 1983; VILLALVA, 1994; PEPERKAMP, 1997; VIGÁRIO, 2001; ALCÂNTARA, 2003; SCHWINDT, 2013, 2014; ULRICH & SCHWINDT, 2018; SANDALO, 2021; entre outros). Para alguns autores, os termos raiz e radical podem dizer respeito a constituintes não isomórficos, sendo a raiz o elemento irreduzível. Neste trabalho, utilizamos os termos indistintamente. Está, a rigor, em jogo aqui a hipótese de que uma base que prescinde de vogal terminal (seja temática seja indicadora de gênero) é a base mínima da derivação. Essa base em geral coincide com o que se entende por raiz ou radical primário, ainda que algumas vezes possa representar uma etapa da derivação que já conta com afixo (ex. atômico(o)+idade → atomicidade). Nesses casos, ainda que a base semântica evidentemente seja a palavra, entendemos que

⁷ Na hipótese de se admitir que o sufixo, ou variante dele, é -aria, *fruto* e *fruta* seriam ambas bases possíveis de *frutaria*. A boa formação, nesse caso, dependeria apenas da queda da vogal final. Mantivemos, contudo, a título de revisão, a lógica do argumento associado ao exemplo sustentada por Matzenauer & Bisol (2016, p. 348).

o afixo se alinha — nos termos de uma restrição de alinhamento generalizado (MCCARTHY & PRINCE, 1993) — ao radical, ignorando a vogal final.

Nesse sentido, destacamos aqui os principais argumentos de Harris (1983), para o espanhol, e de Schwindt (2013), para o português brasileiro. Para o espanhol, Harris (1983) — pioneiro entre todas as análises aqui mencionadas — cita uma série de exemplos que corroboram a falta de necessidade de armazenamento do tema no léxico, entre eles:

(i) a ausência do formativo vocálico em vocábulos com sufixos iniciados por consoante (ex. bueno → bondád / *bonodád);

(ii) a falta de motivação independente para a síncope;

(iii) a inoperância da síncope em formações com “marcadores exóticos” (ex. virus → viral/ *viral).⁸

Com relação ao português brasileiro, Schwindt (2013) associa o comportamento prosódico e morfológico dos afixos, admitindo que a união de um afixo a uma base prosódica pode se dar por processos de incorporação, adjunção ou composição.⁹ No caso da borda da direita da palavra, sufixos podem sofrer incorporação prosódica (ex. lev- + -eza), quando a base é uma estrutura fonológica correspondente à raiz morfológica; ou composição prosódica (ex. leve + -mente), quando as duas estruturas bem formadas apresentam autonomia acentual. Essa hipótese prosódica, por consequência, estipula diferentes bases morfológicas, já que, para ocorrer incorporação, a vogal temática não pode estar presente na formação.

A ideia de que a maior parte dos sufixos se ligam à raiz em português é sustentada por Schwindt a partir dos argumentos de Harris (1987). Considerando-se os argumentos do autor, podemos descrever o problema em português nos termos que seguem.

⁸ Usamos o termo *marcadores exóticos* (e tratamos sua queda como um processo de síncope) no sentido adotado por Harris (1992, p. 66, 75): para designar segmentos terminais incomuns à língua, neste caso, ao português.

⁹ Para Schwindt (2013) e Schwindt & Bisol (2022), o processo de adjunção prosódica envolve a relação de fronteira entre um afixo (uma sílaba ou um par de sílabas não escandidos em palavra) e uma PWD. Porque parece acontecer apenas na borda esquerda da palavra em PB, escapa ao tópico deste artigo.

i) A língua prefere ditongos a hiatos. Contudo, as formas derivadas do português, assim como as do espanhol, formam hiatos, em oposição a manter a palavra em sua configuração silábica original.

(4)	herói	*heroiCina ¹⁰	heroína
	canoa	*canoaCeiro	canoeiro

ii) No PB, há também palavras formadas pelos chamados marcadores exóticos, que têm sua derivação a partir da eliminação da sequência sonora final.

(5)	vírus	*viral	viral
	óculos	*oculosar	ocular

iii) Como terceiro ponto, a língua apresenta casos de derivação regressiva, processo que forma substantivos a partir de uma raiz e o acréscimo da vogal temática.

(6)	ameaçar	*ameaçação	ameaça
	imprimir	*imprimição	impressão

iv) Por fim, palavras derivadas por /l/ + sufixo parecem apresentar evidência para a derivação com base na raiz, como se atesta em Schwindt (2021) em análise de pseudopalavras formadas pelos sufixos -eiro e -ismo (ex. beleira, gortelismo). Os participantes avaliam a consoante /l/ precedente ao afixo como parte da base, e não como consoante epentética, pois, em uma pseudopalavra

¹⁰ A letra C maiúscula representa qualquer consoante da língua, já que não podemos definir com precisão qual seria a consoante presente na forma hipotética. Contudo, Canfield (2010) e Pires (2016) mostram que o português brasileiro apresenta uma tendência de epêntese de /z/ em formas atemáticas (ex. pa-z-ada) e /r/ em formas temáticas (ex. bicha-r-ada).

como *valeiro*, há a preferência por supor que a base pluralizada seja *vales*, e não *vaus*.

Admitir a raiz como base mínima para a derivação não significa dizer que níveis maiores, como o tema e a palavra, não sejam aceitos como bases. O tema é requerido, por exemplo, em nominalizações a partir de verbos (ex. alinh[a]-mento, fornec[i]-mento). Além disso, há sufixos que se unem diretamente à palavra (SCHWINDT, 2013, 2014; ULRICH & SCHWINDT, 2018; ULRICH, 2021) — são os chamados sufixos posicionais, a saber, -inho/-zinho, -íssimo e -mente. Os avaliativos -inho e -zinho aparecem ao lado de bases já flexionadas, como *porquinhos*, *aneizinhos*, *leõezinhos*, reconhecidas pela qualidade vocálica, pela presença da semivogal ou pela qualidade do ditongo nasal, respectivamente (CAMARA JR., 1975; MENUZZI, 1993; MIRANDA, 2002). Palavras formadas pelo avaliativo -íssimo também exibem marcas flexionais prévias à anexação do sufixo, como a qualidade vocálica em *novíssima* (MORENO, 1977). Ainda, o sufixo -mente se une a palavras femininas, como *orgulhosamente* ou *valentonamente* (BASILIO, 1998).

Em vista do exposto, um dos caminhos para falsear a hipótese de morfologia baseada no tema/palavra *versus* na raiz é o exame pormenorizado dos contextos de emergência de uma vogal em formas derivadas, em especial diante de sufixos iniciados por vogal. É o que propomos nesta análise de caráter descritivo.

3 Metodologia

A amostra utilizada neste estudo se constitui de 2.264 adjetivos e substantivos derivados pelos sufixos -al e -oso, extraídos do Dicionário Aurélio Eletrônico, versão 2011.

A variável resposta deste estudo é a presença *versus* a ausência da vogal interveniente, aqui designada VI (ex. *bestial vs. canal*; *precioso vs. chuvoso*). Tomando-se como referência sua presença, a pergunta, assim, que fundamenta a análise estatística é: *qual a contribuição das variáveis em estudo para a*

emergência ou não de uma vogal entre a raiz e os sufixos -al e -oso em itens dicionarizados do PB?

As variáveis investigadas como previsoras neste estudo são relacionadas e exemplificadas no quadro a seguir.

Quadro 1 – Variáveis e fatores analisados

Variável	Fatores	Exemplos
segmento final da base	e/i o/u a consoante temática atemáticas em geral	celestial / volumoso batalmal / caldoso medicinal / argiloso floral / horroroso abacaxizal / nodoso
ponto de articulação da última consoante da base	labial coronal dorsal	verbal / seboso crucial / atencioso musical / terroso
presença de consoante epentética	com epêntese sem epêntese	cafezal / edematoso quinzenal / honroso
número de sílabas da base	até 2 3 4 ou mais	final / brilhoso salival / trabalhoso consonantal / ligamentoso
sufixo	-al -oso	braçal abominoso
alomorfia da base	com alomorfia sem alomorfia	vital / arenoso carvalhal / gostoso
frequência lexical da palavra derivada	log 0,6931472–12,944666	social / religioso (>9,0) ameixoal / caldoso (<1,0)
datação	séc. IX–XV séc. XVI-XVIII	vinhal (séc. IX) sequioso (séc. XVI)

	séc. XIX-XXI sem data	incremental (séc. XXI) cadencioso
--	--------------------------	--------------------------------------

Fonte: Elaborado pelos autores.

Os dados foram listados e classificados quanto às categorias acima em planilha do software Excel e analisados por meio de estatística descritiva na Plataforma R (R CORE TEAM, 2022), interface RStudio (RSTUDIO TEAM, 2022). Os gráficos foram gerados em ggplot2 (WICKHAM, 2009).

4 Resultados e discussão

Nesta seção, partimos da aplicação geral do fenômeno e, em seguida, apresentamos o resultado para cada uma das variáveis relacionadas na seção anterior em termos de distribuições de frequências, acompanhado de discussão.

4.1 Aplicação geral

Apesar de nossos dados registrarem 22,35% (506/2.264) de palavras contendo a estrutura V+al e V+oso (ex. adverbial, ganancioso), apenas 6,14% (139/2.264) são de vogais que podem ser consideradas intervenientes nos termos deste estudo, isto é, vogais que não pertencem à base originalmente e nem ao sufixo (ex. bestial, laborioso). A análise que apresentamos na sequência diz respeito a esse último conjunto de dados. Tratamos, portanto, neste texto, de um fenômeno de muito baixa aplicação, mas cuja descrição é relevante para o mapeamento do domínio do processo de derivação sufixal no idioma.¹¹

¹¹ Os dados foram submetidos a testes de regressão logística de efeitos fixos pela Plataforma R, com uso da função glm. Por se tratar de um conjunto muito restrito de dados, apesar de o modelo apresentar convergência, não se mostrou robusto o suficiente para permitir inferências sobre o fenômeno, de acordo com o índice de determinação R²TJur, inferior a 10% em variados recortes (sem interações e com diferentes tipos de interações).

4.2 Segmento final da base

A maioria dos casos em que emerge uma VI em nossa amostra diz respeito a bases fechadas por *o/u*, em seguida pelas fechadas por C (temáticas), por *a* e, por último, por *e/i*. Não há casos de VI com palavras atemáticas. A Tabela 1 traz o resultado para essa variável.

Tabela 1 – % segmento final da base e VI

	e/i	o/u	a	C temática	atemáticas
com	0,71	3,09	0,97	1,37	0,00
sem	9,10	38,25	32,29	11,84	2,39

($X^2(4) = 29,69$; $p = 5,656e-6$)

Fonte: Elaborado pelos autores.

Embora com números tímidos, o resultado da Tabela 1 poderia sugerir que *o/u* é mais recorrente porque coincide com a vogal final da base, a VT -*o* (ou marca de gênero masculino, para alguns).¹² Essa análise, que favorece a interpretação de sufixação baseada no tema/palavra, contudo, é falseada, em primeiro lugar, pelos resultados obtidos para a vogal *a* na base (já que nunca a VI é *a*; estando bloqueada a coincidência entre vogal da base e VI neste caso, portanto). Além disso, também depõe contra a interpretação de que *o/u* é mais recorrente a eventual divergência entre as supostas VT's -*e/-i* e -*o/-u* e as sugeridas VI's correspondentes, respectivamente *i* e *u*, o que se vê na Tabela 2.

¹² Diferentes combinações foram testadas na organização em se tratando de segmento final da base. Para fins deste artigo, optamos por apresentar uma análise que distingue as vogais átonas, nosso principal foco, de consoantes (em que incluímos ditongos nasais) e de palavras atemáticas. Para fins do exercício de regressão logística que empreendemos, porém, associamos consoantes a palavras atemáticas, unindo os ditongos nasais às vogais átonas, classificados de acordo com a articulação do glide. Em todos os cenários observa-se o efeito, para a emergência de VI, positivo de -*u* e negativo de -*a* e de bases atemáticas, se comparadas essas às demais variáveis.

Tabela 2 – % correspondência segmento final da base e VI

	e/i	o/u	a	C temática	atemáticas
com i interveniente	10,07	6,47	14,39	20,14	NA
com u interveniente	1,44	43,88	1,44	2,16	NA

(N=139; $X^2(3)=82,49$; $p=2,2e-16$)

Fonte: Elaborado pelos autores.

Nesse caso, apesar de predominar a relação *o/u* – *u*, chama atenção o emprego da VI *i* para bases não fechadas por *e/i* (totalizando 41%) e da VI *u* para bases não fechadas por *o/u* (totalizando 5,04%). Aqui, não é a quantidade de casos que tem maior peso de argumento, mas a simples emergência dessas vogais, que por si só fragiliza uma hipótese generalizada de ligação do sufixo direto ao tema/palavra.

4.3 Ponto de articulação da última consoante da base

Em relação ao ponto de articulação da última consoante da base, apesar de se tratar de variável não significativa, observa-se prevalência de coronais no contexto precedente a VI's, como mostra a Tabela 3. As únicas palavras com contexto precedente labial são *canavial*, *grameal* e *nuvioso* e, com contexto precedente dorsal, *arcoal*, *gincoal* e *sequioso*.

Tabela 3 – % ponto de articulação C da base e VI

	labial	coronal	dorsal
com	0,13	5,87	0,13
sem	11,98	69,11	12,77

($X^2(2)= 33,82$; $p=4,536e-8$)

Fonte: Elaborado pelos autores.

Examinadas as palavras contendo coronais como último segmento da base seguidas de VI's+sufixo, observa-se grande predominância das consoantes /t/ e

/s/, as mais recorrentes no final de raízes portuguesas de modo geral. Isso, somado ao fato de que a proporção entre os três tipos de consoantes é muito semelhante também nos contextos *sem VI*, sugere que estamos diante de fenômeno lexical (talvez em termos de frequência fonotática) e não necessariamente de fenômeno assimilatório.

4.4 Consoante epentética

A Tabela 4, a seguir, mostra o comportamento de palavras em que se atesta consoante epentética em relação à emergência de VI. A variável não foi considerada significativa em nossa análise, haja vista escassez generalizada de consoantes epentéticas na amostra (118 itens apenas). De todo modo, a concomitância de consoante epentética com VI para ainda mais excepcional, restringindo-se às palavras *canavial*, *canzoal*, *voluptuoso*.¹³

Tabela 4 – % consoante epentética e VI

	com epêntese	sem epêntese
com	0,13	6,01
sem	5,08	88,78

(X²(1)= 2,18; p=0,1402)

Fonte: Elaborado pelos autores.

Sendo a consoante epentética elemento que emerge para acomodar fonologicamente estruturas morfológicas que se unem e, supondo-se que a VI não corresponde ao segmento final da base, mas tem função semelhante à da consoante epentética, é de se esperar que não co-ocorram.

¹³ A rigor, à exceção de *canzoal*, que em princípio diz respeito à estrutura *cão+z+o+al*, nas palavras *voluptuoso* e *canavial* o que designamos como consoante epentética pode ter uma explicação diacrônica. A primeira se deriva de *voluptas*, do latim, ou seja, já possuía o *t* em sua origem, convertido em ditongo na forma básica da palavra *volúpia* em português. A segunda, há quem defenda, se deriva por analogia com o nome que designa a plantação de *cânave* (ou *cannabis*), planta que, diferente de *cana*, já contém a consoante labial em sua forma básica, ou, ainda, à associação de *cana+via*.

4.5 Número de sílabas da palavra base

A associação de uma VI faz, em princípio, a palavra aumentar em uma sílaba, formando um hiato com o sufixo iniciado por vogal (ex. ra.ça → ra.c[i].+al]; gra.ça → gra.c[i].o.so). Investigar a relação entre VI's e número de sílabas se justifica, assim, para se examinar tanto a evitação a estruturas mais longas quanto a evitação ao hiato – restrições falseadas na literatura sobre a fonologia de diferentes línguas do mundo. Isso se confirma em nosso fenômeno em primeiro lugar pela escassez, em termos de aplicação geral, de formações Raiz+VI+Vsufixo em comparação a estruturas Raiz+Vsufixo, mas também por bases extensas apresentarem menos VI's do que bases curtas, como mostra a Tabela 5. A evitação ao hiato em casos de VI (e consequente redução/manutenção do número de sílabas) é, ainda, alcançada nesses casos pela conversão dos hiatos em ditongos (ex. ra.c[jaw]; gra. c[jo].so).

Tabela 5 – % número de sílabas da base e V interveniente

	até 2	3	4 ou mais
com	2,74	2,56	0,84
sem	28,14	37,15	28,58

($X^2(2) = 21,68$; $p = 1,961e-5$)

Fonte: Elaborado pelos autores.

A hipótese sobre evitação de palavras mais extensas deve ser falseada na perspectiva da relação entre as variáveis *número de sílabas* e *sufixo*, já que -al é monossilábico e -oso é dissilábico. Isso é explorado na próxima subseção.¹⁴

¹⁴ A hipótese de morfologia sufixal baseada na raiz coloca em xeque a relevância da variável *número de sílabas da palavra-base* aqui investigada, porque raízes, em princípio, não estão sujeitas a silabificação, uma propriedade da palavra fonológica. De todo modo, numa abordagem de restrições, entendemos que acessar raiz (em termos de alinhamento a um domínio) é expediente que não entra em contradição com admitir a existência da palavra, com todos os seus atributos.

4.6 Sufixo

A amostra apresenta 1.387 palavras derivadas por -al e 877 palavras derivadas por -oso. Um teste de qui-quadrado aplicado a essa distribuição mostrou não haver relação significativa entre o sufixo e a presença ou ausência de uma vogal interveniente.

Tabela 6 – % sufixo e V interveniente

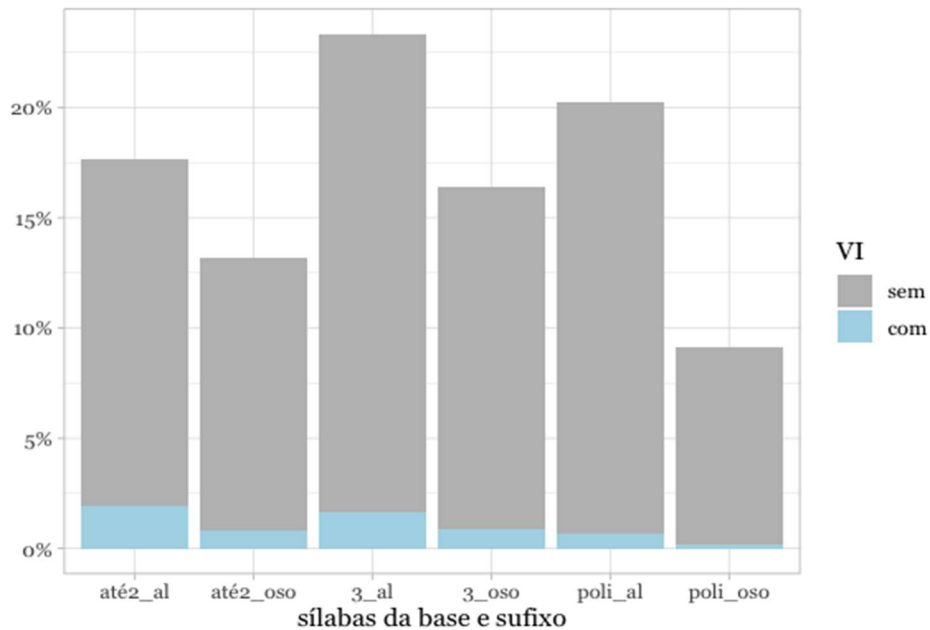
	-al	-oso
com	4,24	1,90
sem	57,02	36,84

($X^2(1) = 3,46$; $p = 0,06303$)

Fonte: Elaborado pelos autores.

Quando cruzada com número de sílabas, porém, a variável *sufixo* se mostra significativa ($X^2(5) = 29,09$; $p = 2,226e-05$). O Gráfico 2, a seguir, mostra a combinação entre cada um dos três padrões silábicos e os dois sufixos investigados.

Gráfico 2 – Sílabas da base & sufixo e VI



Fonte: Elaborado pelos autores.

Como se vê no Gráfico 2, produtos mais curtos, os formados por -al, são preferidos em todos os contextos aos produtos mais longos, os formados por -oso. Um pós-teste aplicado a este cruzamento, porém, mostra que a diferença significativa atestada para este grupo de fatores, no âmbito de VI's, se deve às variáveis *sufixo -al com bases de até 2 sílabas e sufixos -al e -oso com bases polissilábicas*. Apenas como especulação, sugerimos que a motivação esteja novamente na constituição do léxico.

4.7 Base alomórfica

A maior parte dos itens dicionarizados do PB investigados não apresenta base alomórfica. No contexto de VI, essa variável não se mostra significativa.

Tabela 7 – % base alomórfica e V interveniente

	alomórfica	não alomórfica
com	0,71	5,43
sem	7,91	85,95

($X^2(1) = 1,21; p = 0,271$)

Fonte: Elaborado pelos autores.

Por não se tratar de um fenómeno do interior da base, a emergência de VI's não deve em princípio estar relacionada à alomorfia da raiz. Contudo, quando essa alomorfia presente na base envolve hiatos ou ditongos próximos à sua fronteira direita, não se descarta que entre em cena algum tipo de bloqueio de sequências de encontros vocálicos.¹⁵ É possivelmente o caso de bases que se superficializam em sua forma primitiva com -ção em sua borda direita e que, diante do sufixo -al, produzem a sequência -cional (ex. nação → nacional/*nacional, organização → organizacion+al / *organizacional).

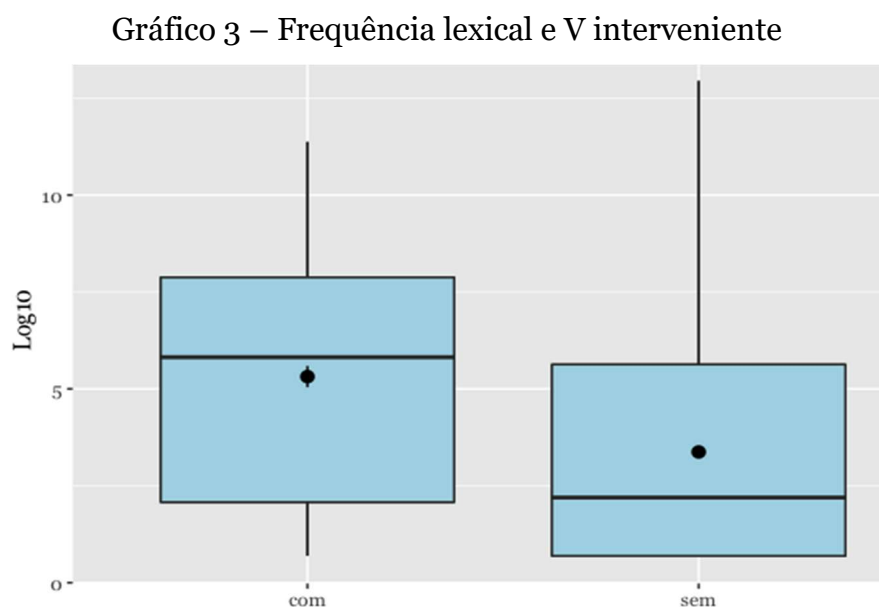
4.8 Frequência lexical da palavra derivada

Os dados de nosso estudo, extraídos do Dicionário Aurélio Eletrônico, foram codificados quanto à frequência, tendo por base o Corpus Brasileiro.¹⁶ O teste Shapiro-Wilk indicou tratar-se de uma distribuição não paramétrica ($W = 0,8509; p < 2,2e-16$). Constituiu-se uma segunda amostra em que se retiraram os *outliers*, a fim de se verificar se essa característica se modificaria, mas ela se confirmou. Tanto para a amostra com *outliers* quanto para a amostra sem *outliers* testou-se, então, a variável numérica *frequência lexical* em duas versões: em valores absolutos e em escala logarítmica. O teste Wilcoxon apontou

¹⁵ Bloqueio semelhante foi defendido por Silveira & Schwindt (2016), no tratamento da competição entre os sufixos -ção e -mento em PB. O que se designou como *evitação de eco* responderia pela não emergência de formas como **acionação* (em favor de *acionamento*) e **amamentamento* (em favor de *amamentação*), em que se registram sequências próximas, respectivamente, da sibilante [s] (cio/ção, na forma básica ção/ção) e da cadeia fonotática [mẽnt] (menta/mento).

¹⁶ <https://www.linguateca.pt/aceso/corpus.php?corpus=CBRAS>.

diferença significativa entre as medianas dos valores absolutos em todas essas versões. Por razões descritivas, optamos, então, por apresentar os resultados em escala logarítmica, como se vê no Gráfico 3, a seguir ($W=87630$; $p<2,2e-16$, para a amostra com \log_{10}).



Fonte: Elaborado pelos autores.

Ainda que palavras sem VI alcancem os índices mais altos de frequência em nossa amostra, considerada a distribuição, itens com VI podem ser considerados mais frequentes. O limite inferior do segundo quartil dos itens com VI se alinha à média dos itens sem VI. Além disso, a distribuição dos itens menos frequentes com VI em muito se aproxima da distribuição dos itens mais frequentes sem VI, se comparados, respectivamente, o segundo quartil do box à esquerda ao terceiro quartil do box à direita do Gráfico 3. A diferença entre as medianas é significativa.

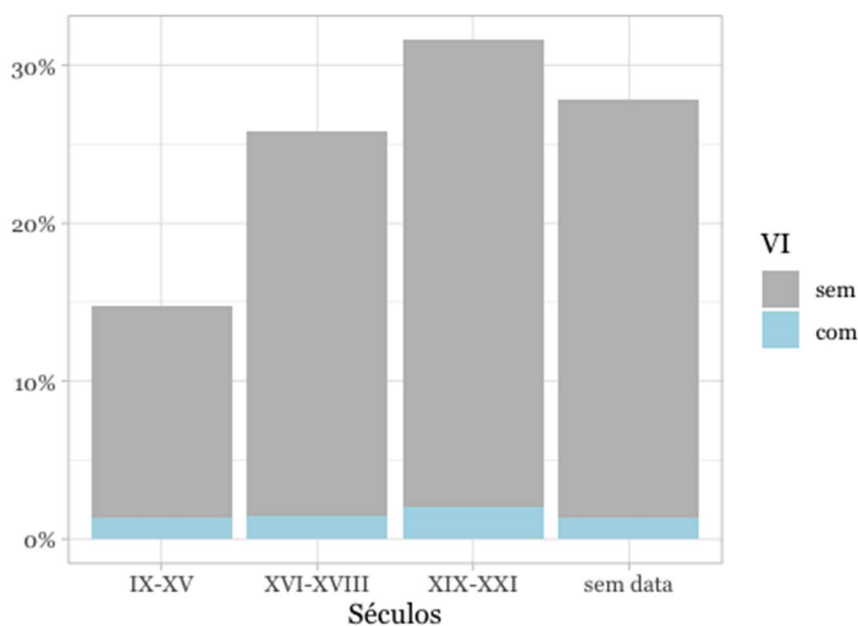
Do ponto de vista da morfologia flexional, alguns autores, entre eles Ullman (1999) e Michel et al. (2011), defendem que formas irregulares são mais frequentes nas línguas do que formas regulares. *Mutatis mutandis*, isso poderia figurar entre as explicações para o fenômeno derivacional aqui analisado (já que lidamos com dois sufixos que, à semelhança de afixos flexionais, são muito produtivos em português). Se considerarmos formas com VI mais irregulares do

que formas sem VI, a distribuição observada estaria justificada. Isso pode se explicar porque estruturas menos deduzíveis estruturalmente exigem mais da memória, e repeti-las mais pode dar conta dessa demanda.

4.9 Datação

Fenômenos escassos em uma língua, como a presença de VI's em formações sufixais, são em geral relacionados à antiguidade (no sentido de idade, tempo de vida) das palavras no léxico. No Gráfico 4, nossos dados estão organizados em 3 faixas temporais, que correspondem à Idade Média (séc. IX-XV), Moderna (séc. XVI-XVIII) e Contemporânea (séc. XIX-XXI), respectivamente, acompanhadas de um conjunto para os quais não localizamos datação, rotulado como *sem data*.

Gráfico 4 – Datação dos vocábulos e V interveniente



Fonte: Elaborado pelos autores.

A distribuição apresentada no Gráfico 4 é significativa ($X^2(3)=8,08$; $p=0,04442$). Examinado o papel individual de cada variável neste grupo, contudo, observa-se que essa significância se deve ao contraste entre o período

séc. IX-XV em relação aos demais, os quais não apresentam diferenças significativas entre si. Pode-se afirmar, então, que palavras mais antigas, que entraram na língua até o século XV, estão mais sujeitas em nossos dados à ocorrência de VI, resultado que contribui para a ideia de que se trata de processo pouco ou não produtivo na língua.

Por fim, em amostra na qual se excluíram dados sem datação (menos 630 dados), comparamos médias e medianas do log de frequência por século. Confirmou-se o observado para essas variáveis em separado: palavras mais antigas, que ingressaram na língua durante a Idade Média, são também as mais frequentes na amostra sob ambas as medidas. As palavras que ingressaram nas Idades Moderna e Contemporânea apresentam, por outro lado, índices de frequência muito próximos.

Tabela 8 – Datação por século e log de frequência lexical

	média	mediana
séc. IX-XV	4.947902	5.007246
séc. XVI-XVIII	3.746152	3.044522
séc. XIX-XXI	3.783032	2.890372
(N=1.634; $X^2=789,71$; $p=0,00884$)		

Fonte: Elaborado pelos autores.

O teste Kruskal-Wallis apontou para uma diferença significativa entre os séculos em relação ao log de frequência. Aplicado o teste Wilcoxon (com método de ajuste Bonferroni) para exame dos pares, constatou-se que a diferença é significativa, a rigor, apenas entre a variável séc. IX-XV e as demais (com valores p , respectivamente, de $1,1e-07$ e $6,8e-08$), mas não entre as variáveis séc. XVI-XVIII e XIX-XXI ($p=1$).

5 Considerações finais

Neste texto apresentamos um estudo sobre a emergência de vogais intervenientes em palavras derivadas pelos sufixos -al e -oso em português brasileiro (ex. best-i-al, grac-i-oso). Tendo como pano de fundo a controvérsia morfológica baseada no tema/palavra *versus* morfológica baseada na raiz, procuramos descrever quantitativamente os contextos de surgimento dessa vogal. Para isso, analisamos todos os substantivos e adjetivos fechados por esses sufixos listados no Dicionário Aurélio Eletrônico (2.264 itens) quanto a variáveis linguísticas. Os principais resultados estão resumidos a seguir.

A aplicação do fenômeno é bastante reduzida na língua, de 6,14%. Em termos de frequência, observou-se que:

i) nem sempre há coincidência entre a vogal final da palavra e a vogal que surge entre raiz e sufixo (ex. graça/gracioso);

ii) a última consoante da base não se mostra relevante para a ocorrência do processo, visto que a aparente preferência por coronais espelha não mais do que uma preferência fonotática geral da língua;

iii) vogais intervenientes e consoantes epentéticas tendem a não ocorrer nos mesmos itens lexicais;

iv) o surgimento de vogais intervenientes acontece mais em bases menores, visto que o acréscimo dessa vogal aumenta a palavra em uma sílaba;

v) há mais vogais intervenientes em formações com -al do que com -oso, padrão possivelmente explicável também pela preferência por aumentar menos a palavra, já que -al é um sufixo monossilábico e -oso é dissilábico;

vi) não parece haver relação entre a ocorrência de VIs e a alomorfa da base;

vii) itens lexicais contendo VIs são mais frequentes na língua, o que pode ser reflexo de uma tendência mais ou menos geral que se atesta, sobretudo na flexão, em formas menos regulares nas línguas;

viii) a proporção do surgimento de palavras com VI entre raiz e sufixo é maior entre os séculos IX-XV, período mais antigo analisado no cópulo — resultado que confirma sua baixa ou nula produtividade no estágio atual da língua.

O reduzido número de ocorrências de palavras com vogais intervenientes na língua refletidos nesta amostra inviabiliza uma análise preditiva robusta do fenômeno analisado. Os resultados da análise frequentista tampouco permitem se identificarem predominâncias importantes de ordem estrutural sobre o processo. Os achados, por outro lado, sugerem se tratar de um fenômeno lexical resultante da história das palavras afetadas pelo processo, que se refletem na constituição do vocabulário *per se* (por ex. na correlação entre número de sílabas da base e número de sílabas do sufixo) e em frequência de uso. A ocorrência de vogais intervenientes entre raiz e sufixo, portanto, não serve para sustentar a hipótese de morfologia sufixal baseada no tema/palavra nem para contrariar a hipótese de morfologia sufixal baseada na raiz.

Referências

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

BASÍLIO, Margarida. Morfológica e Castilhamente: um Estudo das construções X-mente no Português do Brasil. *Delta*, v. 14, São Paulo, 1998.

BERMÚDEZ-OTERO, Ricardo. Morphological structure and phonological domains in Spanish denominal derivation. In: MARTÍNEZ-GIL, Fernando; COLINA, Sonia (eds.) *Optimality-theoretic studies in Spanish phonology*. Amsterdam: John Benjamins, p. 278-311, 2006.

BERMÚDEZ-OTERO, Ricardo. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *Probus*, v. 25 (1), p. 3 - 103. 2013.

BISOL, Leda. O acento e o pé métrico binário. *Caderno de Estudos Linguísticos*, v. 22, p. 68-80. Campinas, 1992.

CÂMARA JR., Joaquim Mattoso. *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão. 2014 [1975]. 158 p.

CAMARA JR, Joaquim Mattoso. *Para o estudo da fonêmica portuguesa*. 2ª ed., Rio de Janeiro: Padrão, 1977.

CANFIELD, Samanta Sá. *A epêntese consonantal no português: um estudo introdutório*. Dissertação (Mestrado em Letras) – PUCRS, Porto Alegre, RS, 2010.

FERREIRA, Aurélio Buarque de Holanda. *Novo Dicionário Eletrônico Aurélio versão 5.11a*. 5. ed. Curitiba: Positivo, 2010. (versão eletrônica)

HARRIS, James. *Syllable structure and stress in Spanish: a nonlinear analysis*. Cambridge: MIT, 1983. 158 f.

HARRIS, James. The form classes of Spanish substantives. In: BOOIJ, Geert; VAN MARLEN, Jaap (eds.) *Yearbook of Morphology 1991*. Dordrecht: Kluwer, p. 65-88, 1992.

HALLE, Morris; MARANTZ, Alec. Distributed morphology and the pieces of inflection. In: HALE, Ken; KEYSER, Samuel Jay. *The View from Building 20*, Cambridge, Mass.: MIT Press, p. 111 - 176, 1993.

KIPARSKY, Paul. Opacity and cyclicity. *The Linguistic Review* 17 (2-4), 351-65, 2000.

LEE, Seung-Hwa. O acento primário do português do Brasil. *Revista de Estudos da Linguagem*, v. 6, n. 2, p. 5-30, Belo Horizonte, 1997.

MATZENAUER, Carmen Lúcia Barreto; BISOL, Leda. O inventário e a distribuição subjacente das vogais temáticas na classe dos nomes do português. *Alfa*, São Paulo, v. 60 (2), p. 341-353, 2016.

MCCARTHY, J.; PRINCE, A. Generalized Alignment: the prosodic morphology interface. In: BOOIJ, G. & MARLA, J. van (eds.). *Yearbook of Morphology 1993*, p.79-153. Dordrecht: Kluwer, 1993.

MENUZZI, Sérgio. *On the Prosody of the Diminutive Alternation -inho/-zinho in Brazilian Portuguese*. Manuscript. HIL/Leiden University, 1993.

MICHEL, Jean-Baptiste *et al.* Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, ed. 331, n.6014, p. 176-182, 2011.

MIRANDA, Ana Ruth. A metafonia nominal (português do Brasil). *Letras de Hoje*, v. 37, n. 1, p. 69-98. Porto Alegre: PUCRS. 2002.

MORENO, Cláudio. *Os Diminutivos em -inho e -zinho e a Delimitação do Vocábulo Nominal no Português*. Dissertação (Mestrado em Letras), Porto Alegre, UFRGS, 1977.

PEPERKAMP, Sharon. Prosodic words. HIL dissertation 34. *The Hague: Holland Academic Graphics*. 1997. 231 f.

- PIRES, Caroline de Castro. *Epêntese consonantal em contexto de juntura morfológica: considerações sobre o sufixo -ada*. Dissertação (Mestrado em Letras), Porto Alegre, PUCRS, 2016.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. Disponível em <https://www.R-project.org/>. Acesso em: 28 fev. 2023.
- RSTUDIO TEAM. *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA, 2022. Disponível em: <http://www.rstudio.com/>. Acesso em: 28 fev. 2023.
- SANDALO, Filomena. Theme vowels and a root-based approach to morphology: evidence from Brazilian Portuguese. *Caderno de Squibs: Temas em estudos formais da linguagem*, v. 7, n. 2, 2021, p. 16-31.
- SCALISE, Sergio. *Generative Morphology*. Dordrecht: Foris, 1984. 237 f.
- SCHWINDT, Luiz Carlos. Palavra fonológica e derivação em português brasileiro: considerações para a arquitetura da gramática. In: BISOL, Leda; COLLISCHONN, Gisela (org.) *Fonologia: teorias e perspectivas*. 1ª ed. Porto Alegre: EDIPUCRS, p. 15 - 28, 2013a.
- SCHWINDT, Luiz Carlos. Prosodic Word and Morphological Derivation in Brazilian Portuguese. Supplemental Proceedings Phonology 2013 (extended abstract). *Linguistic Society of America*, p. 1 - 6, 2014.
- SCHWINDT, Luiz Carlos. Predizibilidade da marcação de gênero em substantivos no português brasileiro. In: Danniell Carvalho; Dorothy Brito. (Org.). *Gênero e língua(gem): formas e usos*. 1ed., Salvador / Bahia: Editora da UFBA, v. 1, p. 279-294, 2020.
- SCHWINDT, Luiz Carlos. Underlying representation of [w]-final words in Brazilian Portuguese: evidence from morphological derivation. *Acta Linguistica Academica*. Budapest: Akadémiai Kiadó, 2021.
- SCHWINDT, Luiz Carlos; BISOL, Leda. The prosodization of derived words: joining perspectives. *Delta*, 28(3), p. 1-27, 2022.
- SILVEIRA, Luciana Morales; SCHWINDT, Luiz Carlos. Alternância do uso de -ção e -mento em nominalizações no português do sul do Brasil. *ReVEL*, edição especial, n. 13, 2016.
- ULLMAN, Michael Thomas. Acceptability Ratings of Regular and Irregular Past-tense Forms: Evidence for a Dual-system Model of Language from Word Frequency and Phonological Neighbourhood Effects. *Language and Cognitive processes*, ed. 14, n.1, p. 47-67, 1999.

ULRICH, Camila Witt. *O estatuto morfológico e prosódico da sufixação avaliativa em português brasileiro*. Tese (Doutorado em Estudos da Linguagem). Universidade Federal do Rio Grande do Sul, 2021. 276 f.

ULRICH, Camila Witt.; SCHWINDT, Luiz Carlos. O status morfoprosódico dos sufixos -inho/-zinho, -mente e -íssimo no português brasileiro. *Delta*, v. 34, n. 2, p. 769 - 788, 2018. Disponível em <https://revistas.pucsp.br/index.php/delta/article/view/39419/26676>.

VIGÁRIO, Marina. *The prosodic word in European Portuguese*. Tese (Doutorado em Linguística). Universidade de Lisboa, 2001. 440 f.

VILLALVA, Alina. *Estruturas Morfológicas: Unidades e Hierarquias nas Palavras do Português*. Tese de doutoramento. Universidade de Lisboa, 1994. 399 f.

WICKHAM, Hadley. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd Edition, Springer, New York, 2009. Disponível em <https://doi.org/10.1007/978-0-387-98141-3>

[Artigo recebido em 28 de fevereiro de 2023 e aceito em 20 de julho de 2023.]