



Frequência e distribuição de plurais irregulares no Corpus Brasileiro

Frequency and distribution of irregular plurals in the Corpus Brasileiro

Luiz Carlos Schwindt

Universidade Federal do Rio Grande do Sul (UFRGS, CNPq), Porto Alegre,
Rio Grande do Sul / Brasil

schwindt@ufrgs.br

<https://orcid.org/0000-0003-0533-589X>

Pedro Eugênio Gaggiola

Universidade Federal do Rio Grande do Sul (UFRGS, PROBIC-FAPERGS),
Porto Alegre, Rio Grande do Sul / Brasil

pedro.e.gaggiola@gmail.com

<https://orcid.org/0000-0002-0123-6205>

Isabela Prisco Petry

Universidade Federal do Rio Grande do Sul (UFRGS, PIBIC-CNPq), Porto Alegre,
Rio Grande do Sul / Brasil

isabelappetry@gmail.com

<https://orcid.org/0000-0003-1082-6589>

Resumo: Neste texto aborda-se a frequência e a distribuição de formas de plural irregular do português brasileiro, no âmbito da palavra, numa perspectiva descritiva. Os dados provêm do Corpus Brasileiro e estão divididos em duas amostras: Amostra L, nomes pluralizados terminados ortograficamente em vogal+is (ex. papéis), vogal+us (ex. chapéus) e is (ex. funis), e Amostra N, os terminados por ões (ex. vilões), ãos (ex. irmãos) e ães (ex. pães). O exame das variáveis fonético-fonológicas e léxico-morfológicas –

número de sílabas, acento, contexto fonológico, afiliação morfológica e frequência lexical – permitiu contextualizar o comportamento das alternantes minoritárias de cada amostra, em oposição às alternantes prevalentes, vogal+is e ões, respectivamente.

Palavras-chave: plural; alomorfia; morfologia; morfologia; morfologia; Corpus Brasileiro.

Abstract: This paper addresses the frequency and distribution of Brazilian Portuguese irregular plurals, within the scope of the word, in a descriptive approach. The data come from the Corpus Brasileiro and are divided into two samples: (i) pluralized nouns ending, in spelling, with vowel+is (eg *papéis* ‘papers’), vowel+us (eg *chapéus* ‘hats’), and is (eg *funis* ‘funnels’), and (ii) those ending with ões (eg *vilões* ‘villains’), ãos (eg *irmãos* ‘brothers’), and ães (eg *cães* ‘dogs’). The phonological and lexical-morphological variables analyzed – number of syllables, stress, phonological context, morphological affiliation and lexical frequency – allowed to define the main contexts for the minority alternants of each sample, in opposition to the prevalent ones, vowel+is and ões, respectively.

Keywords: plural; allomorphy; morphology; morphophonology; Corpus Brasileiro.

Submetido em 09 de outubro de 2020

Aceito em 09 de novembro de 2020

1 Introdução

Neste artigo, tratamos da frequência e da distribuição das realizações de plural de nomes do português brasileiro (PB) terminados, em sua forma singular, nas sequências ortográficas vogal+u/l e ão, a partir de dados do Corpus Brasileiro.¹

Por conveniência metodológica, as alternantes de plural são abordadas neste estudo a partir de uma tipologia que considera as rimas das sílabas que comportam a informação morfológica de plural (isto é, a porção fonológica que inclui o núcleo silábico e todos os segmentos que o sucedem). São **Vis**, **Vus** e **is**, para vogal+u/l, e **ões**, **ãos** e **ães**, para ão, como se exemplifica em (1) e (2), respectivamente.

¹ <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>

(1)	alternante	plural		singular	
a.	Vis	pedais	[pe'dajs]	pedal	[pe'daw] ~ [pe'daʔ]
b.	Vus	chapéus	[ʃa'pews]	chapéu	[ʃa'pew]
c.	is	funis	[fu'nis]	funil	[fu'niw] ~ [fu'niʔ]

(2)	alternante	plural		singular	
a.	ões	balões	[ba'lõʃs]	balão	[ba'lãw̃] ²
b.	ãos	irmãos	[ir'mãws]	irmão	[ir'mãw̃]
c.	ães	capitães	[kapi'tãjs]	capitão	[kapi'tãw̃]

Trata-se de uma abordagem essencialmente descritiva com potencial para contribuir, a partir da observação do léxico em uso, para o debate sobre o papel da produtividade na definição das restrições fonológicas e morfológicas que concorrem para seleção dessas formas não canônicas de plural na língua, as quais tratamos aqui como irregulares.

O comportamento de plurais irregulares em PB foi objeto de muitos estudos, em diferentes perspectivas, em especial no que diz respeito à origem do glide, nos dois casos mencionados, e à representação da nasalidade, no caso dos ditongos em *ão* (CAMARA JR., 1969, 1970; ABAURRE GNERRE, 1983; BISOL, 1998, 2016, 2020; WETZELS,

² Optamos neste texto, como estratégia de simplificação, por nos referirmos aos ditongos em análise por sua representação ortográfica. Nos exemplos de (2), no que diz respeito especificamente aos casos de *ão* e seus plurais, adotamos uma representação fonética que considera uma sequência de vogal e glide nasalizados. Sabemos, contudo, que a questão não é tácita entre fonólogos e foneticistas. Alguns estudiosos defendem a realização da consoante nasal em coda, plena ou secundária, seguindo esse glide. O português parece situar-se, em relação à pronúncia dessa nasal, numa posição intermediária, se considerarmos, num extremo, línguas como o inglês ou o espanhol, que no mais das vezes realizam a nasal plenamente em coda silábica (ex. bu[m] 'bum'; co[n], 'com'), e, noutro, línguas como o francês, que parece exibir uma assimilação completa da nasalidade (ex. av[ã] 'avant'). Esse fato contribuiria para a defesa da hipótese de articulação secundária na língua (ex. irmãw̃ŋ). A consoante nasal é, como discutimos neste texto, subjacente para muitos autores e funcionaria como gatilho para a nasalização do ditongo, em princípio oral na origem. Não sendo apagada, isto é, preservando-se plena ou secundariamente em coda, após o espraçamento, torna-se alvo da assimilação do ponto de articulação do segmento que a sucede ou mesmo do glide que a precede, a depender do contexto.

1997, 2000; GUIMARÃES; NEVINS, 2013; entre outros), mas também na perspectiva à qual se soma esta análise, a da produtividade (HUBACK, 2010a, 2010b; CRISTÓFARO-SILVA, 2012; BECKER *et al.*, 2018; GOMES; PRADO; AMARAL, 2021, no prelo; RIZZATO, 2018; entre outros).

A hipótese geral assumida aqui é a de que restrições implicadas na seleção de expoentes fonológicos concorrentes para entidades morfológicas podem ser alcançadas pela medição da produtividade de seus contextos fonológicos e morfológicos, sobretudo em distribuições imperfeitas, aquelas que não se enquadram plenamente no conceito tradicional de alomorfa. O exercício empírico que descrevemos neste artigo se dá na esfera de hipóteses subordinadas a essa, tratando de cada contexto potencialmente revelador dessas restrições, considerando-se um corpus de grande abrangência.

O texto está organizado como segue. Na seção 2, problematizamos a noção de alomorfa no que diz respeito à realização das marcas de plural exploradas neste trabalho e a relacionamos ao conceito de produtividade que adotamos. Na seção 3, resumimos brevemente alguns estudos que se ligam mais diretamente aos objetivos de nossa pesquisa. Na seção 4, detalhamos os procedimentos metodológicos. Por fim, na seção 5, apresentamos e discutimos os resultados da investigação. Seguem-se nossas considerações finais acompanhadas da agenda da pesquisa.

2 Alomorfa, produtividade e plurais irregulares do PB

Alomorfe, ou variante de morfema, define-se, na tradição, como uma das realizações de um morfema. Nem toda realização de morfema, contudo, enquadra-se nesse conceito. Segundo Bonet, Lloret e Mascaró (2015, p. 1), alomorfa existe sob duas condições: que haja mais de um morfe para cada morfema e que a divergência entre os morfes não possa ser predita pela fonologia regular da língua. Assim, cada alomorfe, por ser imprevisível fonologicamente, constitui uma forma subjacente distinta. Por outro lado, alomorfes não ocorrem irrestritamente: além de, para muitos, deverem se assemelhar fonologicamente, devem também ter especificados seus contextos de ocorrência, e a expectativa é que tais contextos estejam numa relação de excludência mútua, ou distribuição complementar (BAUER, 2004, p. 15).

Nesses termos, podemos nos perguntar se as formas de plural aqui abordadas, exemplificadas em (1) e (2), podem ser consideradas alomorfes.

Para responder a essa pergunta, em primeiro lugar, precisamos examinar se é possível relacionar as formas em análise a um morfema comum, de maior abrangência na língua, no caso -s, além de nos assegurar de que o que as diferencia foneticamente não se caracteriza como regra fonológica ordinária na língua. Começando pelo segundo critério, um simples teste posicional pode responder à questão. Sabemos que as sequências VI ou Vu podem aparecer no meio da palavra (ex. alto, incauto). Não há, contudo, evidência de processo que converta VI em **Vis** no interior do vocábulo, e a sequência ls é estranha ao português (a palavra *solstício* seria uma exceção). **Vus**, por sua vez, parece fazer parte da representação lexical de palavras também em posição interna (ex. claustro, Fausto). A sequência ão, por outro lado, não ocorre, em princípio, em sílaba medial em português, mas apenas em final de palavra. Esse teste indica que os processos que analisamos são licenciados apenas na formação do plural, o que sugere que não estamos diante de fenômenos fonológicos regulares da língua. A complexidade do problema, porém, reside na tarefa de se explicar, por via fonológica, alternantes que modificam a forma básica de plural -s ou que lhe acrescentam substância fônica. Nessa perspectiva, várias hipóteses são exploradas na literatura, a maioria defendendo que está em jogo, mais do que alomorfia de plural, alomorfia de raiz (ex. papele+s/ limon+s) e emergência de um glide ativado pela associação de [s] à sílaba da forma singular (BISOL, 1998; WETZELS, 1997; entre outros). Em contraste está a alternativa de se lexicalizar o morfema, mais do que sua base, assumindo-se, se não alomorfia propriamente dita, a hipótese de subléxicos, ativados por restrições fonológicas ou de outra natureza, inclusive extralinguística. É, por exemplo, a análise assumida por Becker *et al.* (2018) e Rizzatto (2018), para o tratamento de plurais de palavras fechadas foneticamente por Vw, ou por Abaurre Gnerre (1983), para o tratamento de plurais de palavras fechadas por ditongo nasal.

Em segundo lugar, precisamos considerar que os contextos de ocorrência dessas formas não são cem por cento excludentes, ou seja, em princípio, as três variantes ocorrem em ambientes fonéticos e gramaticais semelhantes. Há, porém, predominâncias em relação a esses contextos que podem sinalizar para quase-regularidades.

No recorte da análise desenvolvido neste texto, vamos nos fixar nessas sub-regularidades, sem compromisso com o status mais ou menos alomórfico das alternantes que analisamos, uma vez que as propriedades fonológicas das marcas de plural, ou mesmo sua afiliação morfológica – à base ou ao sufixo –, são por ora secundárias.

Assumimos o entendimento de que produtividade de morfemas é propriedade gramatical das línguas naturais. Isso quer dizer que tanto pode ser depreendida do léxico em uso quanto pode ser predita a partir de restrições que concorrem para formação de novos itens, o léxico potencial. O léxico potencial, portanto, deve dar conta da porção transparente do léxico disponível, mas também pode incluir formações menos regulares, fruto de subconjuntos de restrições ou de restrições mais baixas na hierarquia de uma língua. Este estudo dedica-se a dados do léxico em uso.

3 Estudos sobre a produtividade de plurais irregulares em PB

A alternância dos expoentes fonológicos do plural de ditongos orais e nasais foi avaliada na perspectiva de sua produtividade em contextos fonológicos e morfológicos específicos por Huback (2010a, 2010b), Cristófaró-Silva (2012), Becker *et al.* (2018), Rizzato (2018), Gomes, Prado e Amaral (2021, no prelo), entre outros autores. Nesta seção apresentamos algumas das principais ideias desses textos em versão resumida. Por conveniência expositiva, tratamos primeiramente dos plurais em u/l, para, em seguida, tratarmos dos plurais em ão.

Plurais em u/l

Em relação a palavras terminadas em u/l, Huback (2010a) realiza um estudo envolvendo a aplicação de um teste de reação a 36 falantes nativos do PB, contando com 53 palavras-alvo distribuídas em 3 categorias relativas a frequência lexical. Com base na amostra do Corpus NILC/São Carlos, as palavras-alvo são classificadas nas frequências de ocorrência baixa, média e alta. Com foco teórico no Modelo de Redes (BYBEE, 2001), está em jogo a hesitação ou não hesitação do falante ao produzir o plural. Os resultados mostraram que 12,8% dos itens terminados em l foram pluralizados como se pertencessem aos itens terminados em u, ou seja, contando somente com a adição da expressão fonológica de plural -s. A autora observou favorecimento em particular

da hesitação para os itens *mel*, *sol* e *sal*. Apesar de haver diferenças de frequência entre esses vocábulos, o que os aproxima é o fato de serem todos monossílabos, sugerindo papel desta variável em favor da hesitação. De modo geral, a autora conclui que não há correlação entre frequência de ocorrência e hesitação, já que palavras de frequência de ocorrência baixa, média e alta inibiram hesitações. Os itens terminados em u mostraram-se mais suscetíveis a hesitações, o que pode ser explicado pelo fato de esse ser o grupo com menor frequência de tipo. Desses itens, 15,6% migraram para a pluralização em is. Itens de baixa frequência, como *jirau*, *mausoléu*, *véu*, apresentaram índices mais altos de hesitação, confirmando as expectativas do modelo adotado. O item *judéu*, contudo, de frequência alta no corpus adotado, contrariou a expectativa. As palavras menos suscetíveis a hesitação nesse grupo foram justamente as de frequência mais alta, como *meu* e *seu*, apontando para correlação entre as variáveis frequência de ocorrência e hesitação no grupo dos itens fechados por u. O estudo conclui, ainda, que a frequência de tipo, menor para os itens terminados em u, também interfere nas migrações, mais recorrentes nesse grupo.

Ainda a respeito de palavras terminadas em u/l, Cristófaros-Silva (2012), propondo uma análise à luz do que refere como modelos multirrepresentacionais, categoriza as marcas fonológicas de plural desses itens em (i) nomes em que a marca de plural ocorre a partir da adição de -s (ex. *degraus*) e em (ii) nomes em que o plural se realiza pela adição de um suposto morfema *is* (ex. *sais*), com alteração do radical (além de formas em que a marca de plural pode deixar de ocorrer, como em *os sal*). A autora destaca, com base nos dados de Huback (2007), que é possível se atestarem plurais pertencentes ao padrão (i) realizando-se conforme o padrão (ii) (ex. *degrais*).

Becker *et al.* (2018) apresentam um estudo sobre essas formas na perspectiva da aquisição de seus plurais. Os autores realizam um experimento psicolinguístico envolvendo pseudopalavras com 115 crianças de 7 a 12 anos e um grupo de controle formado por adultos. Participantes de 7 a 9 anos evitaram a alternância [w ~ j] em monossílabos, sugerindo influência de uma restrição da língua que atua na proteção da primeira sílaba, o que, na interpretação dos autores, equivale a proteção de monossílabos. O estudo mostra que essa proteção é ainda maior entre crianças de 10 a 12 anos. O grupo de controle e as crianças de 10 a 12 anos, além disso, tendem a evitar novos ditongos com baixa dispersão

de altura entre vogal e glide (ex. [ej, ew, oj, ow]), apesar de sua forte presença no léxico. Esse fato também é interpretado na perspectiva de uma restrição violável, que milita pela maior dispersão de altura em ditongos.

Gomes, Prado e Amaral (2021, no prelo) também realizam um experimento psicolinguístico valendo-se de pseudopalavras, além de um teste de produção com palavras de baixa frequência. Participaram 54 voluntários, sendo 25 com ensino superior, cursando a graduação da Faculdade de Letras na UFRJ, e 28 de um curso de Educação para Jovens e Adultos de Niterói, a fim de controlar o papel da escolaridade. Os resultados apontam para prevalência da forma de plural *s* em monossílabos para ambos os níveis de escolarização. Em relação a essa variável, entretanto, constataram divergência quanto à vogal [e] no núcleo da última sílaba tônica, que se mostrou significativa apenas para o grupo de participantes de nível superior, desfavorecendo o plural *js*.

Plurais em ão

No âmbito do ditongo nasal, Huback (2010a) verifica também a disponibilidade de *ões*, *ãos* e *ães* no léxico dos informantes quando perguntados a respeito do plural de vocábulos como *escrivão*, por exemplo. Os resultados obtidos a partir do experimento, descrito anteriormente, apontam para uma migração direcionada no uso das alternantes de plural do ditongo nasal: 32,5% dos vocábulos cujo étimo prevê *ãos* e 20,9% dos vocábulos cujo étimo prevê *ães* foram pluralizados pela alternante *ões* no experimento. Há, portanto, influência da frequência de tipo da alternante *ões* em sua produtividade, visto que *ões* participa da pluralização de um maior grupo de palavras terminadas pelo ditongo nasal em PB de acordo com dados do dicionário Houaiss observados por Huback (2010a, p. 19). A migração no sentido contrário não foi atestada de maneira expressiva: 4,1% de itens etimologicamente pluralizados pela alternante *ões* optaram por *ãos* para expressar plural, por exemplo. Houve, entretanto, hesitação nas respostas obtidas no experimento. Na análise de regressão binária, vocábulos para os quais se supõe *ãos* no étimo e vocábulos de baixa frequência de ocorrência se mostraram favorecedores de hesitação na obtenção de seus plurais, apontando para influência significativa de efeitos de frequência de tipo e de ocorrência no fenômeno em questão. Huback (2010b) apresenta resultados semelhantes a partir de outro experimento, incluindo agora a leitura de frases e figuras

como estímulos, considerando fatores linguísticos e extralinguísticos. Em relação ao número de sílabas, por exemplo, constatou que monossílabos desfavorecem categoricamente a aplicação do expoente fonológico mais produtivo, *ões*.

Cristófar-Silva (2012) também discute a organização do plural de nomes do PB terminados em ditongo nasal na perspectiva de modelos multirrepresentacionais. A autora classifica nomes encerrados pelo ditongo nasal *ãos* na categoria geral de nomes que expressam plural pelo simples acréscimo de *-s* (ex. irmãos). Vocábulos que tomam *ões* ou *ães* como expoente fonológico de plural (ex. leões, pães), por outro lado, se relacionam pelo fato de apresentarem, segundo a autora, *alteração no radical nominal* e por compartilharem o morfema de plural *-is*. O padrão do grupo de itens lexicais que toma *ões* como marca de plural é aplicado em generalizações, de acordo com a análise, como consequência de sua alta frequência de tipo, em consonância com a análise de Huback (2010a).

Essa preferência pelo expoente *ões* é percebida também por Rizzato (2018), que promove uma análise experimental, aplicada a 79 falantes do PB, envolvendo a seleção de plural de ditongos nasais em 24 frases. O contexto analisado é o do aumentativo *-zão* – expressão de grau passível de anexação em vocábulos que já contenham traço de número (ex. coraçõezões), em sua interpretação. A opção pela marcação dupla de plural se revela variável em seus dados, ainda que a preferência por *ões* no âmbito do sufixo seja significativa. Embora sob variação, monossílabos exibem um comportamento distinto também nesta análise, sendo *pãezões* e *cãezões* as formas preferidas, enquanto trissílabos compartilhadores do mesmo plural etimológico optam pela expressão de plural menos redundante (ex. capitãozões).

4 Procedimentos metodológicos

Nesta seção apresentamos a constituição e a organização de duas subamostras extraídas a partir de uma amostra-base de 3.744.513 *types* e 691.758.151 *tokens* disponível para download no site do Corpus Brasileiro – doravante CBras. O CBras é um banco de dados alimentado por diferentes fontes, incluindo fala e escrita. A análise que empreendemos neste estudo tem como base principalmente os *types* oferecidos pela amostra-base; recorreremos, contudo, aos *tokens* na discussão sobre frequência lexical.

As subamostras analisadas estão assim constituídas:

- (i) Amostra L, contendo 9.245 vocábulos pluralizados correspondentes a bases fechadas ortograficamente por vogal+u/l;
- (ii) Amostra N, contendo 5.899 vocábulos pluralizados correspondentes a bases fechadas ortograficamente por ão.

Os dados foram extraídos da amostra-base e organizados com o uso da Plataforma R. A investigação restringiu-se ao que tratamos como *nomes*, numa interpretação ampla desse termo, que abriga substantivos e adjetivos – vocábulos sujeitos à flexão de plural. A organização inicial dos dados consistiu, entre outros aspectos, em juntar formas com pequenas diferenças de grafia, incluindo maiúsculas e minúsculas ou espaçamentos indevidos, eliminar resíduos de palavras estrangeiras ou com sequências incompreensíveis e checar no corpus, através de busca pelo site Linguateca,³ a classe gramatical de itens homófonos para excluir com segurança não nomes (ex. vão, não). Os dados, contudo, neste caso, não são analisados em sua forma plenamente lematizada, porque optamos pela manutenção da distinção entre lexemas, isto é, palavras simples, derivadas ou compostas foram preservadas separadamente na amostra. No exercício analítico apresentado neste texto, porém, por diversas vezes apresentamos comparações com subconjuntos de dados que fazem referência a lemas, controlados como uma entre as variáveis independentes.

Após a etapa de preparação automática dos dados, cada subamostra foi codificada manualmente levando-se em conta propriedades fonológicas e léxico-morfológicas que pudessem contribuir em alguma medida para contextualizar a ocorrência das diferentes formas de plural em foco.

Ainda que as análises das duas subamostras estudadas sejam inteiramente independentes uma da outra, por serem conduzidas a partir de variáveis análogas, são apresentadas nesta seção e na seção de resultados sempre que possível conjuntamente.

Não é demais registrar, ainda, que, apesar de usarmos neste texto o termo *variável*, comum em estatística, não abordamos o fenômeno em questão como variável no sentido da sociolinguística variacionista, isto é,

³ <https://www.linguateca.pt/aceso/corpus.php?corpus=CBRAS>

entendemos que estamos tratando de uma alternância, já que os contextos de seleção de cada variante são na maioria das vezes excludentes. Variação, enquanto *duas ou mais formas de se dizer a mesma coisa*, ocorre em margem muito reduzida na amostra (ex. g[ojs] ~ g[ows] ~ g[ols]; capit[õjs] ~ capit[ãws] ~ capit[ãjs]) e não é nosso foco aqui.

4.1 Variável dependente

As variáveis de interesse em nossa análise são as alternantes de plural atestadas em cada amostra, como exemplificamos inicialmente em (1) e (2). A subdivisão que propomos é tão somente uma alternativa de análise, naturalmente passível de reinterpretação a depender de como se defina o padrão de marcação morfológica de plural nesses casos. Como dissemos, optamos nas duas amostras por representar essas variáveis considerando as rimas das sílabas finais das formas pluralizadas.

Alternantes de plural da Amostra L

A divisão adotada inicialmente para a Amostra L contempla três alternantes de plural: **Vis**, **Vus** e **is**. A primeira e a última alternantes, embora se assemelhem no sentido de terminarem em **is**, diferenciam-se quanto à realização obrigatória de um ditongo no primeiro caso (ex. past[ɛjs]) e à não realização ou realização apenas opcional de um ditongo homorgânico, ou ainda de um alongamento da vogal, no segundo (ex. vin[is] ~ vin[ijs] ~ vin[i:s]). A segunda alternante, **Vus**, também forma ditongo, mas com um glide labiovelar (ex. r[ɛws]).

Alternantes de plural da Amostra N

Para a Amostra N, as alternantes em análise são também três, as tipicamente registradas no léxico do português: ões, ãos e ães (ex. bot[õjs], m[ãws], c[ãjs], respectivamente).

4.2 Variáveis independentes

As variáveis independentes deste estudo, examinadas sempre na perspectiva das alternantes mencionadas, são baseadas em hipóteses da literatura resenhada na seção 3 e em aspectos que consideramos não plenamente contemplados nesses estudos. Os grupos de fatores investigados são a seguir descritos e exemplificados.

Número de sílabas

Nas duas amostras, os vocábulos foram classificados quanto ao número de sílabas da palavra. A primeira classificação incluiu palavras de 1 a 9 sílabas. Uma reorganização, entretanto, foi proposta, seja porque dados de 5 a 9 sílabas se mostraram escassos, seja porque não se observou, numa rodada preliminar, relevância na distinção entre palavras com 4 sílabas e palavras maiores. A classificação analisada, assim, foi **1 sílaba** (ex. réus, cães), **2 sílabas** (ex. quartéis, irmãos), **3 sílabas** (ex. mausoléu, capitães), **4 ou mais sílabas/polissílabas** (ex. governamentais, nacionalizações).

Acento

Palavras terminadas em vogal+u/l e ão apresentam acento final na maioria dos casos e, em proporção consideravelmente reduzida, são acentuadas na penúltima sílaba, justificando a seguinte classificação: **acento final** (ex. quintais, sabões) e **acento pré-final** (ex. móveis, órgãos).

Contexto fonológico

Neste grupo foram classificados os segmentos que precedem a porção fonológica comum às três alternantes estudadas em cada amostra. Assim, na Amostra L, abordamos a vogal do núcleo da sílaba envolvida no plural, a que antecede o glide no ditongo (ex. p[a]us, c[ε]us, anz[ɔ]is, vin[i]s, az[u]is). No caso de [i], estão contempladas as possibilidades de não ditongação, de ditongação homorgânica ou de alongamento (ex. vin[i]s ~ vin[ij]s ~ vin[i:]s). Na Amostra N, abordamos a consoante do onset da sílaba que contém o ditongo nasal. Uma classificação inicial considerando todos os segmentos individualmente foi reinterpretada na perspectiva de modo de articulação: **oclusiva** (ex. capi[t]ães), **fricativa** (ex. se[s]ões), **líquida** (ex. sa[l]ões) e **nasal** (ir[m]ãos). Também foram considerados casos de **hiato** (ex. pe.ões).

Afiliação morfológica

Neste grupo de fatores, analisamos a localização morfológica da terminação da base que dá origem aos plurais investigados. Interessa-nos dizer se as porções fonológicas u ou l, na Amostra L, e ão, na Amostra N,

coincidem integralmente com um sufixo (ex. pastor+**il**, mulher+**ão**), são apenas parte de um sufixo (ex. amá+**vel**, implementa+**ção**) ou integram o radical (ex. **mel**, **cão**). Para isso, propomos uma divisão mais acurada de contextos para cada amostra, que se sujeitou a amalgamações resultando, na Amostra L, em radical mais 3 tipos de sufixos, e, na Amostra N, em radical mais 4 tipos de sufixos, como se detalha na seção 5.4.

Frequência lexical

A frequência lexical dos itens pluralizados em análise é a informada na amostra disponibilizada pelo CBras e foi analisada como uma variável contínua.

Outras variáveis, como padrão de **terminação da base**, **classe gramatical**, **lema** e **frequência do lema**, também foram codificadas e aparecem nesta análise como subsidiárias à descrição das 5 que tomamos como nucleares.

5 Resultados e discussão

Nesta seção fazemos uma apresentação dos resultados da análise das duas subamostras investigadas. Como antecipamos, essa exposição se dá, sempre que possível, de forma combinada, ainda que as análises tenham sido computadas independentemente.

Os resultados aqui apresentados são produto de estatística descritiva fazendo-se uso da Plataforma R. Cada resultado é acompanhado de discussão, considerando-se os achados de trabalhos anteriores e o potencial de cada grupo de fatores para uma futura análise de natureza preditiva.

Começamos por apresentar a frequência geral das alternantes de plural para cada uma das amostras analisadas. Na sequência, tratamos das 5 variáveis que consideramos nucleares na ordem em que foram apresentadas na metodologia. As variáveis subsidiárias são chamadas quando necessárias à discussão dos resultados obtidos para as categorias nucleares.

5.1 Frequência geral

Confirmando o amplamente divulgado na literatura, observamos, na Amostra L, prevalência da alternante **Vis**, seguida por **Vus** e **is**

com índices semelhantes, e, na Amostra N, prevalência da alternante **ões**, seguida de **ãos** e **ães** — esta última, contudo, com ocorrência consideravelmente inferior à alternante **ãos**.

TABELA 1 – Distribuição geral de plurais irregulares – Amostras L e N

	Amostra L			Amostra N		
	%	Ocor.		%	Ocor.	
Vis	98,0	9.061	papéis	92,8	5.472	balões
is	1,0	88	funis	5,2	307	irmãos
Vus	1,0	96	chapéus	2	120	capitães
Total		9.245		Total	5.899	

Por conta da distribuição geral apresentada na Tabela 1, em que se opõem, em cada amostra, uma alternante de emprego superior a 90% a outras duas alternantes com emprego bastante reduzido, nossos resultados serão na maioria das vezes apresentados, a partir daqui, considerando-se um universo de 100% para cada alternante. Isso permitirá focalizar o papel das variáveis examinadas no espectro de cada alternante, dando visibilidade às variantes não prevalentes, e também possibilitará comparar esse efeito entre as alternantes em termos relativos.

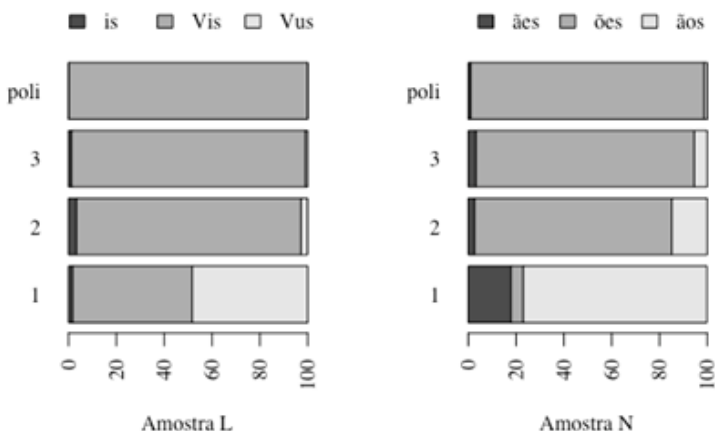
5.2 Número de sílabas

O exame da variável número de sílabas revelou, para as duas amostras investigadas, que, quanto maior a extensão silábica do vocábulo, menor a ocorrência das alternantes não prevalentes.

Na Amostra L, **Vis** é a alternante prevalente em vocábulos de 2 ou mais sílabas. Entre os monossílabos, contudo, divide a liderança com **Vus**, que apresenta 50% de ocorrência. Há apenas um caso, 1,8%, relativo à alternante **is** entre os monossílabos.

Na Amostra N, a alternante **ões** também é prevalente em vocábulos de 2 ou mais sílabas. Entre os monossílabos, contudo, observa-se uma inversão de preferência, com **ãos** como a variante mais frequente, 76,9%, seguida de **ães**, 17,9%, contra reduzidos 5,1% de **ões**. Não devem ser desprezados nesta amostra também os índices encontrados para **ãos** entre dissílabos e trissílabos.

GRÁFICO 1 – % de plurais irregulares e nº de sílabas – Amostras L e N



O comportamento diferenciado de monossílabos relatado por Huback (2010a, 2010b), em especial para plurais de ão, foi atestado em nossas duas amostras. Pode estar em jogo aqui a fidelidade de monossílabos às suas formas de base, conforme defendem Becker, Nevins e Levine (2012) e Becker *et al.* (2018), numa abordagem que parecia monossílabos com sílabas iniciais. Essa tese, contudo, fica na dependência do debate sobre que formas estão disponíveis na base, ou na subjacência, dessas alternantes: se contam com segmentos mais abstratos, como /l/, por exemplo, no caso da Amostra L, e /N/, no caso da Amostra N. Embora tópico de nossa investigação, representações subjacentes não são, contudo, foco deste artigo, razão por que por ora adiamos esse debate. Além disso, considerando que nosso fenômeno se restringe ao contexto de final de palavra, afirmações sobre comportamento análogo de monossílabos a sílabas iniciais seriam, neste recorte, temerárias.

Um exercício necessário, entretanto, é a verificação dos itens lexicais – em nosso caso, lemas – que integram esses padrões excepcionais, a fim de subsidiar a discussão sobre um possível efeito de lexicalização de palavras mais do que de morfemas ou alomorfes neste caso. A hipótese é de que quanto menor a proporção de lemas para ocorrências, maior a restrição lexical, isto é, maior a possibilidade de um item se repetir na amostra. Essa, evidentemente, é uma medida geral, que deve ser relativizada tanto nos casos em que há muito poucas ocorrências

quanto nos casos em que itens particulares apresentam frequências muito desequilibradas em relação aos demais dados.

Na Tabela 2, registramos a distribuição de lemas e monossílabos atestados para as alternantes investigadas em cada amostra devidamente listados/exemplificados (em sua forma pluralizada, por conveniência expositiva). Itens considerados excepcionais estão em destaque. Os números mostram limitação categórica de lemas para a variante **is**, na Amostra L, e para a variante **ões**, na Amostra N. No que diz respeito às demais variantes, as proporções da distribuição lemas/ocorrências, todas entre 20 e 30%, sugerem importante limitação lexical. Esse fato, considerada a frequência individual dos itens, pode ser um indicador de controle do léxico mais do que de seleção alomórfica propriamente dita. Essa discussão será, na perspectiva da frequência lexical geral, retomada na seção 5.6.

TABELA 2 – Plurais de monossílabos – Amostras L e N

		Amostra L			Amostra N		
	%	Lemas/ Ocor.		%	Lemas/ Ocor.		
Vis	29,6	8/27	sais, réis, géis, méis, sóis, grais, gois , móis	ões	100	2/2	chões, vões
Vus	32,1	9/28	graus, maus, réus, céus, paus, naus, véus, vaus, tchaus	ãos	23,3	7/30	mãos, grãos, vãos, sãos, chãos, nãos, pãos
is	100	1/1	vis	ães	28,6	2/7	cães, pães

Entre os dissílabos, trissílabos e polissílabos, apesar da prevalência das variantes **Vis**, na Amostra L, e **ões**, na Amostra N, as alternantes minoritárias também merecem alguma atenção.

Na Amostra L, observa-se, como no caso dos monossílabos, restrição lexical no caso da alternante **Vus**, predominando, entre os dissílabos, os itens *degraus*, *chapéus*, *troféus* e *mingaus*, e, entre os trissílabos, vocábulos como *mausoléus*, *berimbaus* e *bacalhaus*, além de alguns nomes próprios pluralizados, como *nicolaus* ou *venceslaus*. Não há polissílabos no padrão **Vus**. No caso da variante **is**, contudo, o fenômeno parece mais automático, já que não há concentração de lemas, mas uma

relação direta entre a terminação Consoante+il e o plural **is**, tanto em dissílabos, quanto em trissílabos e polissílabos (ex. *civis, perfis, infantis, mercantis, studentis, primaveris*). Exploramos novamente essa relação na seção 5.5, quando abordamos a afiliação morfológica das alternantes em questão, incluindo-se os sufixos.

Na Amostra N, em relação ao padrão **ãos**, também há alguma predominância de certos lemas entre dissílabos, com destaque para os itens *órgãos, irmãos, cristãos, órfãos, bênçãos e pagãos*, e entre trissílabos, concentrando maior frequência em itens como *cidadãos, artesãos, anciãos, acórdãos, afegãos, cortesãos e corrimãos*. Os polissílabos em **ãos** da amostra podem ser considerados formas inesperadas (no sentido de serem muito mais recorrentes com a alternante **ões**): *concentrações, informações, anfitriões, apresentações, comunicações*, entre outras de mesmo tipo, todas com baixa frequência. No padrão **ães**, também se atestam vocábulos de todas as extensões. Entre os dissílabos e polissílabos analisados, todos parecem ser exceções a formas mais comumente realizadas com **ões** e, em alguns casos, com **ãos**, como *refrães, peães, irmães, anães*, entre outros. No que se refere aos trissílabos, diferentemente, entre os itens mais frequentes estão alguns que são predominantemente realizados com **ães** na fala culta, como *alemães, capitães, catalães*, casos muitas vezes de sufixos gentílicos, o que se descreve, como prenunciamos, na seção 5.5, adiante. Os dados de polissílabos, como os de dissílabos, à exceção do vocábulo *tabeliães*, são corriqueiramente produzidos com **ões**, como *cirurgiães, informações, opiniães, alterações* etc.⁴

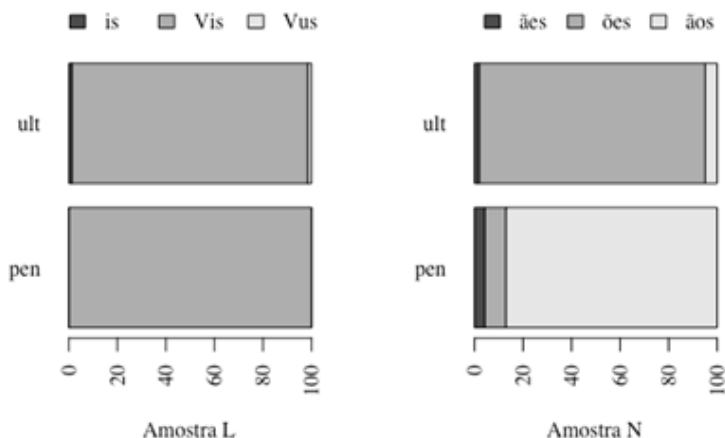
5.3 Acento

Formas terminadas em vogal+u/l e ão ortográficos são predominantemente oxítonas em português. Na Amostra L, 70% dos itens apresentam acento final, contra 30% de paroxítonas. Na Amostra N, quase a totalidade dos itens é de oxítonas, com apenas 0,4% de paroxítonas. Em

⁴ Por se tratar de uma grande base de dados, alimentada por diferentes fontes, não se descarta a possibilidade de esses e outros itens excepcionais atestados nesta análise constituírem meros erros de registro. Não nos cabe, contudo, decidir por descartá-los, tanto porque não apresentam marcas incontroversas de lapsos de grafia quanto porque, apesar de atípicos, apresentam-se como possibilidades na língua. A frequência lexical, por outro lado, figura, em nosso entendimento, como regulador – necessário, se não suficiente – na interpretação dessa ambiguidade.

nenhuma das amostras há proparoxítonas. O Gráfico 2 é particularmente informativo em relação aos dados atípicos.

GRÁFICO 2 – % de plurais irregulares e acento – Amostras L e N



Na Amostra L, 100% das paroxítonas fazem plural em **Vis**. Entre os itens de acento final, contudo, ainda que predomine o padrão **Vis**, 1,5% e 1,3% correspondem, respectivamente, aos padrões **Vus** e **is**. A quantificação das ocorrências e a proporção de lemas envolvidos nesses dois padrões excepcionais estão listados na Tabela 3. Os lemas representam, tanto para a alternante **Vus** quanto para **is**, mais de 50% das ocorrências nessas categorias, índices que não permitem conclusões contundentes acerca de restrição lexical.

TABELA 3 – Plurais de oxítonos – Amostra L

	%	Lemas/Ocor.	
Vus	76	73/96	graus, maus, réus, céus, degraus, chapéus
is	53,5	46/86	civis, perfis, infantis, juvenis, barris, studentis

Na Amostra N, localizamos 23 casos de paroxítonas. Como mostra a Tabela 4, são os mesmos 8 lemas caracteristicamente realizados no plural com a alternante **ãos** que apresentam alguma variação em **ões**

e **ães** – estes últimos com baixa frequência de acordo com os índices do CBras. Se de fato nomes paroxítonos condicionam em alguma medida a seleção da alternante **ãos**, não há que se falar em restrição lexical. O fato, porém, de lidarmos com muito poucas ocorrências neste caso deixa tal conclusão em suspenso, na dependência de análises que considerem o léxico potencial mais do que o institucionalizado.

TABELA 4 – Plurais de paroxítonos – Amostra N

	%	Lemas/Ocor.	
ões	100	2/2	órgões, acórdões
ãos	40	8/20	órgãos, órfãos, bênçãos, acórdãos, sótãos, cóvaos, cristóvãos, orégãos
ães	100	1/1	órgães

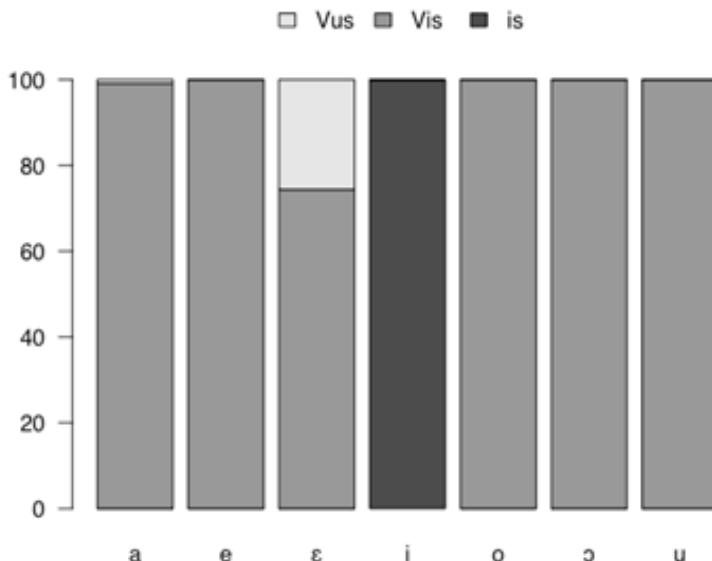
5.4 Contexto fonológico

Nesta subseção tratamos dos segmentos que precedem a porção fonológica comum às três alternantes estudadas em cada amostra. Assim, na Amostra L, abordamos a vogal do núcleo da sílaba envolvida no plural, a que antecede o glide no ditongo ou a que precede s, em casos de suposta não ditongação, como em *funis*. Na Amostra N, abordamos a consoante do onset da sílaba que contém o ditongo nasal, ou sua ausência, em caso de hiato, como em *peões*.

Vogal do núcleo na Amostra L

Na Amostra L, classificamos inicialmente cada uma das vogais que nuclearizam a sílaba envolvida na alternância em questão. A preferência categórica de **Vis** para qualquer vogal é frustrada pela vogal [i], que, por razões estruturais, concentra 100% de suas ocorrências no padrão **is**, e pelas vogais [a] e [ɛ], que apresentam, respectivamente, 0,9% e 25,6% de suas ocorrências no padrão **Vus**.

GRÁFICO 3 – % de plurais irregulares e vogal do núcleo – Amostra L



Os resultados não indicam evidência imediata para agrupar as vogais por algum parâmetro fonético-fonológico (articulatório, por exemplo), razão por que as mantivemos separadas nesta análise. Cabe, contudo, averiguar os itens que justificam esse comportamento não categorial, a fim de checar se não se pode atribuí-lo ao *design* do léxico mais do que propriamente à fonologia. Os dados de [i], categóricos para *is*, serão examinados com mais detalhe quando tratarmos da variável afiliação morfológica, mais especificamente da terminação *il*. Exploramos, porém, a distribuição dos casos atípicos envolvendo as vogais [a] e [ε], como se vê na tabela 5. As proporções relativamente altas de lemas/ocorrências não permitem se concluir que estejamos diante de um efeito de restrição lexical neste caso.

TABELA 5 – Plurais atípicos com núcleos a, ε – Amostra L

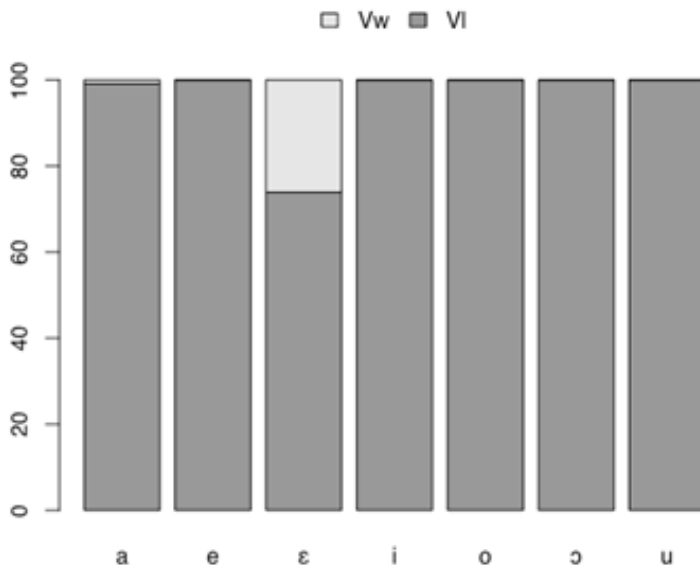
		%	Lemas/Ocor.	
Vus	a	74,5	38/51	maus, degraus, paus, naus, saraus, mingaus
	ε	77,8	35/45	réus, céus, chapéus, troféus, véus, mausoléus ⁵

Desconsiderando-se, contudo, a explicação de restrição lexical, a interpretação para o resultado da alternante **Vus** deve levar em conta, ainda, a hipótese de distinção subjacente entre bases terminadas em /Vw/, ditongos legítimos, e bases terminadas /Vl/, sujeitas à ditongação por derivação fonológica (BISOL, 1989; CAMARA JR., 1970, entre outros). Essa distinção está presente na ortografia inclusive, contrastando pares como *mau* (adjetivo) vs. *mal* (advérbio), cujos plurais ainda se preservam *maus/males* em grande medida na língua culta. Não há garantia de que o dado de escrita, neste caso, reflita a fala, mas também não é possível negar com segurança que alguma representação próxima da escrita possa subjazer, na mente, o dado de fala. O fato é que convenções de escrita podem ser causa e consequência de padrões fonológicos, já que, alimentadas pela fala, são também alimentadoras das representações sonoras mentais de falantes escolarizados (SCHWINDT *et al.*, 2007).

A título de exercício, para explorar especificamente os padrões de terminação da base singular, codificamos adicionalmente a Amostra L da seguinte forma: Vl equivale a formas fechadas na escrita por l (ex. canal, pincel) e Vw, por ditongo com u (ex. pau, céu). O cruzamento dessas categorias com as vogais que ocupam o núcleo silábico está representado no Gráfico 4.

⁵ Desses itens frequentes de base *éu* que listamos, apenas *troféu* apresenta dado de plural com a alternante *Vis*, **troféis**, com frequência 4, contra *troféus*, com frequência 913.

GRÁFICO 4 – % de terminações da base singular e vogal do núcleo – Amostra L



No padrão de base singular VI, que responde por 98,9% dos dados, podemos encontrar todas as 7 vogais como núcleos na Amostra L, concentrando-se 94,5% dos dados nas vogais [a] e [e] (ex. *avental*, *móvel*). No caso de Vw, só se registram dados com núcleos [a] e [ε] (ex. *degrau*, *troféu*), com as ocorrências que já relatamos quando tratamos do padrão de plural **Vus**.

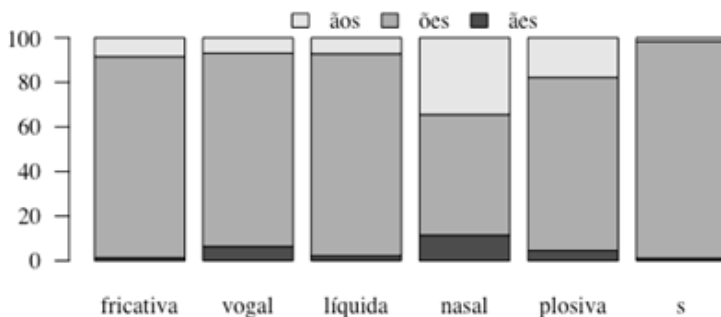
Ainda que não possamos, com base no tipo de dado de que nos utilizamos neste trabalho, concluir sobre a distinção da representação subjacente das alternantes envolvidas, nossos resultados confirmam, por ora, no que tange às vogais nucleares, a correlação entre os padrões VI_{SING}/VI_{PL} , Vw_{SING}/Vw_{PL} .

Consoante precedente na Amostra N

Em relação ao contexto fonológico precedente ao ditongo nasal, a partir do exame estatístico de diferentes classificações considerando os mais variados segmentos e encontros consonantais atestados, chegamos

ao resultado apresentado no gráfico 5. Para todas os contextos, prevalece a alternante **ões**. De diferente da categorização tradicional em grandes classes de sons, nossa classificação inclui a consoante [s], por representar 91,8% das fricativas e 69,6% de todos os contextos.

GRÁFICO 5 – % de plurais irregulares e onset – Amostra N



O comportamento diferenciado de [s], que concentra 97,4% de seus dados na alternante **ões**, coincide com a grande incidência do sufixo **-ção** nos dados. Esse fato é explorado do ponto de vista do sufixo na subseção seguinte. Em relação aos demais contextos não se observam predominâncias que justifiquem em princípio a defesa de uma hipótese assimilatória. Na tabela a seguir, exploramos o comportamento lexical das alternantes não prevalentes em relação à consoante precedente. Os únicos contextos que permitem alguma exploração sobre restrição lexical são, no caso da alternante **ãos**, o das oclusivas e nasais, e, no caso de **ães**, o das nasais. Nesses contextos, a proporção lemas/ocorrências é baixa o suficiente para sugerir alguma medida de restrição lexical.

TABELA 6 – Plurais atípicos e contexto precedente – Amostra N

		%	Lemas/ Ocor.	
	hiato	100	14/14	anciãos, aldeãos, anfitriãos, aviãos
	s	98,4	61/62	infecções, condições, informações, alterações
ãos	fricativo	64,5	20/31	órfãos, artesãos, vãos, cortesãos, chãos
	líquido	62,1	18/29	grãos, refrãos, vilãos, catalãos
	oclusivo	25,6	30/117	órgão, cidadão, cristão, pagão, acórdão
	nasal	14,8	8/54	mãos, irmãos, corrimãos, nãos, alemãos
	hiato	92,3	12/13	guardiães, tabeliães, anciães, cirurgiães
	s	97,8	44/45	infecções, condições, informações, alterações
ães	líquido	88,9	8/9	capelães, catalães, refrães, castelães, tecelães
	fricativo	80	4/5	escrivães, alazães, decisões, lesães
	oclusivo	63,3	19/30	cães, pães, capitães, charlatães, cristães
	nasal	22,2	4/18	alemães, irmães, anães, caimães

5.5 Afiliação morfológica

Como mencionamos, os vocábulos aqui analisados são substantivos e adjetivos, que rotulamos como *nomes*. Considerando-se os propósitos desta etapa de nossa pesquisa, trabalhamos com uma lista de palavras, não de frases, disponibilizada pelo CBras. Ainda que tenhamos quantificado os vocábulos por classe, muitos itens restaram dúbios, já que adjetivos podem facilmente ser empregados como substantivos em português, assim como substantivos podem eventualmente ser empregados como adjetivos.⁶ Decidimos, por outro lado, no âmbito das

⁶ Embora a busca do item dentro da frase seja acessível aos usuários do CBras por meio do site Linguateca, não recorremos a esse expediente para checar se nomes de classe dúbia haviam sido empregados como substantivos ou como adjetivos, por consideramos essa informação, operacionalmente custosa, pouco relevante para os fins de nossa pesquisa. Essa checagem restringiu-se à desambiguação, como mencionamos na metodologia, de nomes com outras classes apenas (como verbos ou advérbios, por ex.).

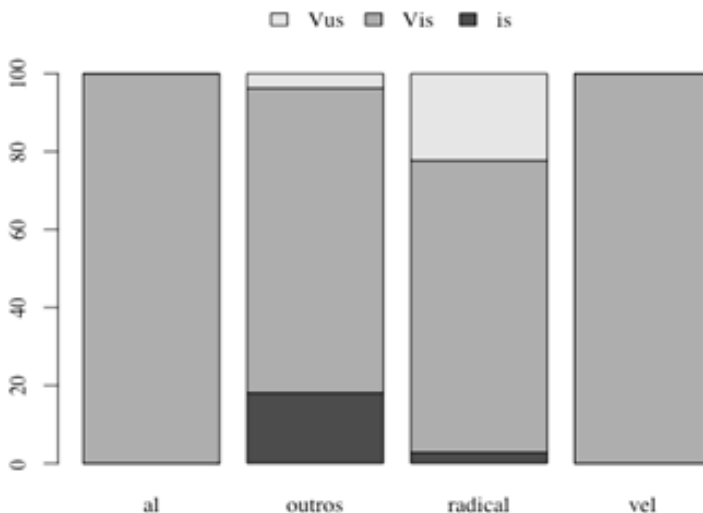
variáveis morfológicas, centrar nossa discussão na afiliação morfológica das alternantes, já que as formas de base para os plurais aqui investigados podem se localizar no radical, coincidir integralmente com sufixos ou, ainda, apenas ser parte de sufixos. Como afixos derivacionais estão relacionados à categorização dos vocábulos, porém, um debate sobre classes pode eventualmente ser desenvolvido a partir do exame desta categoria.

Nesta seção, como na anterior, por conta das peculiaridades de cada amostra no que diz respeito à morfologia, apresentamos seus respectivos resultados separadamente.

Afiliação morfológica na Amostra L

Foram inicialmente examinados 18 contextos, sendo 17 de sufixos potenciais (já que nem todos podem ser seguramente considerados sufixos na sincronia da língua) e formas em que as terminações de base u/l faziam parte do radical (ou raiz acrescida ou não de elementos na borda esquerda). Esses contextos foram reestruturados em apenas 4 categorias, considerando-se que 91,3% dos dados correspondem aos sufixos -al e -vel, distribuindo-se os demais dados entre o radical e outros sufixos comparativamente pouco frequentes.

GRÁFICO 6 – % plurais irregulares e afiliação morfológica da base – Amostra L

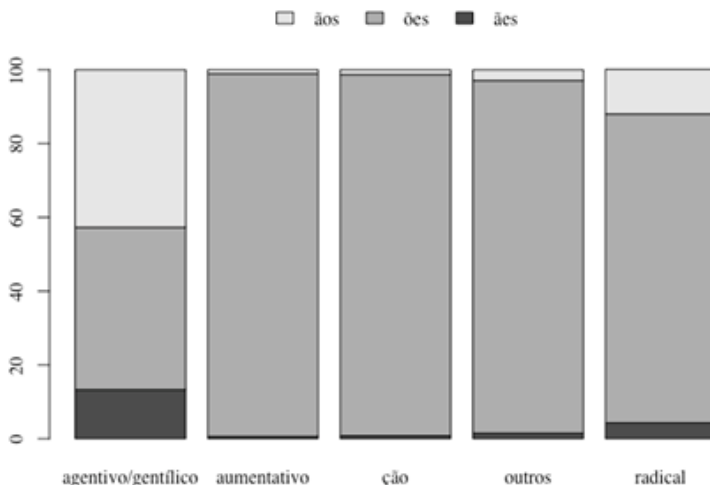


Todos os casos de bases fechadas pelos sufixos -al e -vel fazem plural em **Vis**. Entre outros sufixos e radical, registram-se ocorrências das alternantes **Vus** e **is**. Entre os sufixos que classificamos como *outros* – 4,7% da amostra – estão casos sobretudo de -eu (ex. europeus), εu (ex. fogaréus), -el, (ex. carretéis), -ol (ex. espanhóis) e -il (ex. febris). A incidência destacada da alternante **is** na categoria *outros* deve-se, em primeiro lugar, às ocorrências do sufixo il, bastante produtivo na formação de adjetivos em português (ex. estudantil, febril, juvenil), mas que nem sempre consegue ser claramente isolado de suas bases (ex. civil, sutil, hostil), e, em segundo lugar, a itens em que il é parte da raiz (ex. perfil, fuzil, refil). A variável *radical* acomoda vocábulos em que não se detecta qualquer processo sufixal (ex. paus, sóis) ou aqueles para os quais a opacidade da fronteira sufixal parece consolidada (ex. ramais, quartéis). Por essa razão, os dados da alternante **Vus** se concentram nessa categoria. Por fim, enfatizamos que o índice para a variável *radical* pode se ampliar se consideramos que parte não desprezível das palavras classificadas em *outros sufixos*, por suspeição da transparência sufixal, poderia estar aptas a migrar para essa categoria.

Afiliação morfológica na Amostra N

A partir de uma divisão inicial em 10 contextos, que procurava dar conta de 8 sufixos e 2 tipos de bases, considerando-se a distribuição dos dados, chegamos a uma classificação que contempla 5 contextos apenas, o radical mais 4 tipos de sufixos. Antes de apreciarmos o gráfico a seguir, que tem por base, como vimos adotando, 100% das ocorrências para cada contexto distribuídas entre as alternantes investigadas, é importante se registrar que 61,4% dos dados da Amostra N dizem respeito ao sufixo -ção (ex. construções), 20,2% ao radical (ex. pães) e 10,9 ao aumentativo (ex. lixões). Os 7,5% de dados restantes distribuem-se entre agentivos/gentílicos (ex. artesãos, catalães) e outros sufixos (ex. colisões).

GRÁFICO 7 – % plurais irregulares e afiliação morfológica da base – Amostra N



Os casos de -ção, de aumentativo e de outros sufixos selecionam a alternante **ões** predominantemente, podendo ser enquadradas suas exceções na pequena margem de variação ou mesmo de erro de registro de escrita a que está sujeito o tipo de dado de que nos utilizamos neste trabalho (ex. durações, informações, tristãos, beatães, decisões, opiniões).

No caso dos radicais e dos sufixos agentevos/gentílicos, exploramos hipótese diversa: um pouco mais numerosos, os casos atípicos poderiam se referir a restrições lexicais. É o que se explora na tabela 7, que permite identificar indício de restrição lexical no subconjunto dos sufixos agentevos/gentílicos especialmente no que se refere à alternante **ãos**.

TABELA 7 – Plurais atípicos e afiliação morfológica – Amostra N

		%	Lemas/ Ocor.	
ãos	radical	50,7	73/144	órgãos, mãos, grãos, irmãos, órfãos
	agentevo/gentílico	19,4	20/103	cidadãos, cristãos, artesãos, pagãos
ães	radical	80,4	41/51	cãos, pãos, capitãos, tabeliãos, refrãos
	agentevo/gentílico	46,9	15/32	alemão, guardião, escrivão, capelão

Sobre a relação entre afiliação morfológica e contexto fonológico precedente na Amostra N, cabe apenas um registro sobre a relação -ção/[s]. Trata-se de uma correspondência assimétrica. Obviamente as ocorrências do sufixo -ção (ex. manifestações) coincidem integralmente com as ocorrências de [s]. O contexto precedente [s], porém, corresponde também à maior parte, 37,8%, dos casos de radical (ex. nações), a 4,84% dos casos de sufixo aumentativo (ex. servições) e a 1,66% dos de agentivo/gentílico (todos os casos do lema *saxões*), compreendendo, de modo geral, 8,25% dos dados da Amostra N. Em todos os casos, como vimos, a forma de plural atestada preferencialmente é **ões**. Esses fatos dificultam se afirmar com segurança que o sufixo, e não a consoante [s] (ou sua associação à classe das fricativas), define a seleção da alternante.

5.6 Frequência lexical

Embora tenhamos discutido em alguma medida o papel de itens lexicais específicos ao longo da apresentação dos resultados das variáveis que investigamos, dedicamos essa seção para uma análise mais geral do efeito da frequência lexical sobre a realização das formas de plural em estudo.

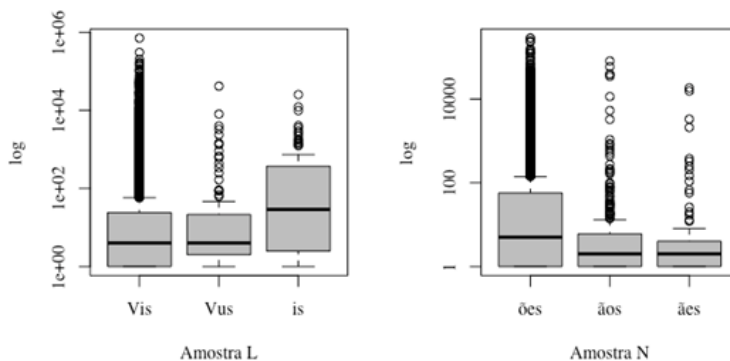
Começamos por listar, com as nuvens de palavras presentes no gráfico 8, os 50 itens mais frequentes de cada amostra.

GRÁFICO 8 – 50 itens de plural irregular mais frequentes – Amostras L e N



No gráfico 9, tratamos da frequência dos itens lexicais com maior detalhamento para cada uma das amostras, fazendo uso de uma escala logarítmica. A motivação para a adoção dessa escala é a discrepância de dados absolutos da variante **Vis** sobre **Vus** e **is**, na Amostra L, e da variante **ões** sobre **ãos** e **ães**, na Amostra N.

GRÁFICO 9 – Plurais irregulares e frequência lexical – Amostras L e N



Analisamos primeiramente os resultados para a Amostra L. Em relação à posição, embora a média de frequência para as três alternantes seja relativamente equilibrada, a despeito da grande diferença de dados, observamos que a mediana se iguala para as alternantes **Vis** e **Vus**, situando-se na frequência 4, mas se eleva a 26,5 para alternante **is**. Apesar disso, é **Vus** que apresenta assimetria positiva, concentrando dados no terceiro quartil. Quanto à dispersão, as variantes apresentam importante diferença de amplitude, que reflete a distribuição da amostra, mas também a grande incidência de *outliers* (dados que se afastam do padrão geral de distribuição) na porção superior dos dados. O intervalo interquartil entre as variantes **Vis** e **Vus**, apesar da grande distinção no número total de dados, se assemelha, mas a alternante **is** destoa importantemente das demais. Por fim, merecem destaque as caudas do limite inferior: o segundo quartil se inicia no limite mínimo de frequência lexical, 1, para a alternante **Vis**, mas em 2 para as variantes **Vus** e **is**. As caudas superiores se assemelham para as três alternantes, e o limite superior é equilibrado para **Vis** e **Vus**, situando-se entre as frequências 50 e 60, mas

é consideravelmente maior para **is**, alcançando 884. A tabela 8 contribui para a melhor leitura desses dados.

TABELA 8 – Plurais irregulares e frequência lexical em quantis – Amostra L

Alternante	Limite inferior	25%	50%	75%	Limite superior	100%	Média
Vis	1	1	4	24	58,5	710.897	722,2
Vus	1	2	4	21,3	50,3	41.745	692,3
is	1	2	26,5	354,8	884	25.373	900.64

Interpretando esses resultados na perspectiva da frequência lexical, podemos dizer que os números obtidos para a Amostra L conferem pouco destaque às alternantes **Vis** e **Vus** no que concerne à frequência lexical, mas também que não diferenciam essas alternantes entre si de modo importante. Isso, associado ao comportamento diferenciado de **is**, que conta com mais itens frequentes do que suas concorrentes na porção central da amostra, sugere algum controle à generalização de **Vis** como marca de plural de palavras terminadas em u/l, apesar de sua indiscutível prevalência na língua.

Analisamos agora os resultados para a Amostra N. No que concerne à posição, se, por um lado, as médias das duas primeiras alternantes, **ões** e **ãos**, não se mostram tão distanciadas no comparativo com a terceira alternante, **ães**, por outro, a mediana afasta a primeira, com frequência lexical 5, das duas últimas, com frequência lexical 2. As três alternantes apresentam algum grau de assimetria positiva, com mais itens no terceiro quartil, com destaque para **ões**, alternante em que o índice se eleva mais nesta porção. Em relação à dispersão, como na Amostra L, há aqui importante diferença de amplitude, com *outliers* na porção superior para as três alternantes, acompanhando os índices de prevalência dessas variantes no uso. O intervalo interquartil também é consideravelmente maior para a variante **ões**. O limite inferior é idêntico para as três variantes, coincidindo o início do segundo quartil com a frequência lexical mínima da amostra, 1. As caudas superiores também não se distinguem importantemente, sendo inclusive proporcionalmente idênticas para a primeira e a terceira variantes, respectivamente a mais e a menos prevalentes da Amostra. O limite superior, contudo, é sensivelmente

distinto para as três variantes, acompanhando o caráter decrescente de emprego dessas alternantes. A tabela 9 detalha esses valores.

TABELA 9 – Plurais irregulares e frequência lexical em quantis – Amostra N

Alternante	Limite inferior	25%	50%	75%	Limite superior	100%	Média
ões	1	1	5	57	141	284.580	1.183,9
ãos	1	1	2	6	13,5	80.948	906,7
ães	1	1	2	4	8,5	18.386	342,8

Os resultados obtidos para a Amostra N não sugerem predominância de itens mais frequentes para as variantes marginais, **ãos** e **ães**. Isso somado ao fato de que a variante prevalente no uso, **ões**, concentra também itens lexicais mais frequentes contribui para a ideia de generalização dessa variante como marca de plural de palavras terminadas em *ão* na língua.

6 Considerações finais

Neste texto apresentamos resultados de um estudo descritivo sobre a expressão fonológica de formas de plural irregular em português, no âmbito da palavra, com base em dados do Corpus Brasileiro. Duas amostras foram consideradas: Amostra L, relativa ao plural de nomes terminados, no singular, em vogal+u/l ortográficos, a que correspondem as alternantes **Vis**, **Vus** e **is**, e Amostra N, relativa ao plural de nomes terminados em *ão* ortográfico, a que correspondem as alternantes **ões**, **ãos** e **ães**. Variáveis fonológicas e léxico-morfológicas foram quantificadas em relação a cada uma dessas alternantes.

O estudo confirmou a prevalência, amplamente relatada na literatura, das variantes **Vis** e **ões**, respectivamente, para as Amostras L e N. Dedicamos, por isso, nossa maior atenção às variantes menos frequentes. Em relação a essas variantes, merecem destaque os aspectos a seguir resumidos.

(i) Em relação ao número de sílabas, monossílabos mostram comportamento diferenciado nas duas amostras, como relatado por Huback (2010a, 2010b); Becker *et al.* (2018), Rizzato (2018), entre

outros. Na Amostra L, o destaque é para **Vus**, que se aproxima muito em emprego da variante predominante, **Vis**. Na Amostra N, destaca-se **ãos**, com grande vantagem em relação às demais alternantes, seguido de **ães**. O contraste entre ocorrências e lemas sugere que os padrões excepcionais encontrados entre os monossílabos são, em grande medida, restritos lexicalmente.

(ii) Quanto ao acento, a grande maioria dos dados é de oxítonas, o que apenas confirma uma hipótese geral sobre a preferência do padrão acentual do português, já que estamos diante de sílabas supostamente pesadas (admitindo-se uma consoante ou mesmo um glide em coda). Não há proparoxítonas. As reduzidas paroxítonas atestadas, no caso da Amostra L, seguem o padrão predominante, selecionando **Vis**, e se referem, majoritariamente a formas sufixadas (sincrônica ou diacronicamente). No caso da Amostra N, as paroxítonas selecionam predominantemente **ãos**, em ocorrências bastante restritas lexicalmente.

(iii) Em relação ao contexto fonológico, examinamos, no caso da Amostra L, a vogal que nucleariza a sílaba envolvida em cada alternante. Para todas as vogais há predomínio da alternante **Vis**, com exceção de [i], que realiza **is**, por restrição estrutural, já que todas as formas dizem respeito à terminação *il*. As vogais [a] e [ɛ] estão presentes em dados com a alternante **Vus**, sem indícios de restrição lexical, dada a alta proporção lema/ocorrência nessas categorias. Esse comportamento das vogais se confirma quando se contrastam padrões de terminação das formas pluralizadas a padrões de terminação das bases. No caso da Amostra N, porque a vogal nuclear, [ã], é comum à forma singular das três alternantes, examinamos a consoante que precede esta vogal. O destaque é para [s], que corresponde a quase 70% dos dados, distribuído entre as três variantes, com privilégio para **ões**. As demais consoantes distribuem-se de modo relativamente equilibrado. Identifica-se controle lexical apenas no caso de oclusivas e nasais precedendo **ãos** e de nasais precedendo **ães**.

(iv) No que concerne à afiliação morfológica das alternantes, na Amostra L, 91,3% dos dados são de palavras terminadas nos sufixos *-al* ou *-vel*, todos realizando a variante **Vis**. Os demais dados distribuem-se entre radical e outros sufixos. Desses, destacam-se os dados do sufixo *-il*, todos realizando **is**, e dos ditongos localizados na raiz, que realizam **Vus**. Na Amostra N, 61,4% dizem respeito ao sufixo *-ção* e 20,2%, ao radical. Os demais itens são fechados por sufixos aumentativos, por agentivos/gentílicos e outros sufixos minoritários. Em todos os casos predomina

a alternante **ões**, à exceção dos agentivos/gentílicos, grupo em que se destacam as alternantes **ãos** e **ães**, com alguma restrição lexical. As alternantes **ãos** e **ães**, embora em número mais reduzido e sem aparente controle lexical, também se destacam no contexto dos radicais.

(v) Por fim, quanto à frequência lexical geral das amostras estudadas, podemos dizer, em relação à Amostra L, que o comportamento similar das alternantes **Vis** e **Vus**, apesar de sua considerável diferença em termos de ocorrências, combinado ao comportamento próprio de **is** podem indicar alguma resistência para a generalização de **Vis** como marca de plural de palavras terminadas em u/l. No caso da Amostra N, a variante mais recorrente, **ões**, é também a que concentra maior frequência lexical, sem destaque para o comportamento de **ãos** e **ães**. Esses fatos contribuem para a ideia de generalização dessa variante como marca de plural de palavras terminadas em **ão** na língua.

Este estudo integra um projeto maior, que trata da representação de plurais irregulares em português brasileiro. Sua contribuição descritiva se dá pelo mapeamento de itens do léxico a partir de um banco com grande volume de dados de uso falado e escrito da língua. Esse mapeamento é ponto de partida da análise experimental em curso, em que discutimos, em perspectiva inferencial, o uso dessas alternantes e, sobretudo, suas representações de base.

Agradecimento

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processos PQ 310921/2018-0 e PIBIC 154093/2020-3, e à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), processo PROBIC 20/2551-0000315-9, pelo fomento em forma de bolsas. Agradecemos aos colegas Pedro Surreaux, Rodrigo Mahfuz e Júlia Ricardo.

Contribuição dos Autores

Este artigo corresponde a um recorte de pesquisa sobre representação de plurais no português brasileiro, concebida e coordenada pelo primeiro autor. Os três autores participaram das etapas de levantamento e análise dos dados, de discussão dos resultados e de redação deste texto.

Referências

- ABAURRE GNERRE, M. B. M. Alguns casos de formação de plural em português: uma abordagem natural. *Cadernos De Estudos Linguísticos*, Campinas, v. 5, p. 127-156, 1983.
- BAUER, L. *A Glossary of Morphology*. Edinburgh: Edinburgh University Press, 2004.
- BECKER, M. *et al.* The Acquisition Path of [w]-final Plurals in Brazilian Portuguese. *Journal of Portuguese Linguistics*, Lisboa, v. 17, n. 4, p. 1-17, 2018. DOI: <https://doi.org/10.5334/jpl.189>. Disponível em: <https://jpl.letras.ulisboa.pt/articles/10.5334/jpl.189/>. Acesso em: 11 set. 2020.
- BECKER, M.; NEVINS, A.; LEVINE, J. Asymmetries in Generalizing to and from Initial Syllables. *Language*, Washington, DC, v. 88, n. 2, p. 231-268, 2012. DOI: <https://doi.org/10.1353/lan.2012.0049>. Disponível em: https://becker.phonologist.org/projects/english/becker_nevins_levine_english_2012.pdf. Acesso em: 11 set. 2020.
- BISOL, L. O ditongo na perspectiva da fonologia atual. *DELTA*, São Paulo, v. 5, n. 2, p. 185-224, 1989.
- BISOL, L. A nasalidade, um velho tema. *DELTA*, São Paulo, v.14, nº especial, p. 27-46, 1998. DOI: <https://doi.org/10.1590/S0102-44501998000300004>. Disponível em: <https://revistas.pucsp.br/delta/article/view/43390/28850>. Acesso em: 11 set. 2020.
- BISOL, L. A nasalidade fonológica no português e suas restrições. *Diadorim*, Rio de Janeiro, v. 18, p. 116-126, 2016. DOI: <https://doi.org/10.35520/diadorim.2016.v18n0a4050>. Disponível em: <https://revistas.ufrj.br/index.php/diadorim/article/view/4050>. Acesso em: 11 set. 2020.
- BISOL, L. Sufixos de duas faces. *Revista da Abralin*, Aracaju, v. 19, n. 1, p. 1-12, 2020. DOI: [HTTPS://doi.org/10.25189/rabralin.v19i1.1380](https://doi.org/10.25189/rabralin.v19i1.1380). Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1380>. Acesso em: 11 set. 2020.
- BONET, E.; LLORET, M. R.; MASCARÓ, J. The Prenominal Allomorphy Syndrome. In: _____. (org.). *Understanding Allomorphy*. Perspectives from Optimality Theory. Bristol: Equinox Publishing, 2015. v. 5, p. 1-44.

BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001. DOI: <https://doi.org/10.1017/CBO9780511612886>

CAMARA JR., J. M. *Problemas de Lingüística Descritiva*. Petrópolis: Editora Vozes, 1969.

CAMARA JR., J. M. *Estrutura da Língua Portuguesa*. 35. ed. Rio de Janeiro: Editora Vozes, 1970.

CRISTÓFARO-SILVA, T. Organização fonológica de marcas de plural no português brasileiro: uma abordagem multirrepresentacional. *Revista da Abralin*, Curitiba, v. 11, p. 273-305, 2012. DOI: 10.5380/rabl.v11i1.32468. Disponível em: <https://revistas.ufpr.br/abralin/article/view/32468>. Acesso em: 11 set. 2020.

GOMES, C. A., PRADO, L. O. do; AMARAL, T. L. A. Aspectos cognitivos e sociais da variação linguística na alternância de formas de plural de nomes do PB. In: ORSINI, M.; CAVALCANTE, S. R.; MARINS, J. (org.). *Contribuições à descrição e ao ensino do português brasileiro: da fonética ao discurso, com parada obrigatória na sintaxe (título provisório)*. Rio de Janeiro: EDUF RJ, 2021. No prelo.

GUIMARÃES, M.; NEVINS, A. Probing the Representation of Nasal Vowel in Brazilian Portuguese with Language Games. *ORGANON*, Porto Alegre, v. 28, n. 54. p. 155-178, 2013. DOI: <https://doi.org/10.22456/2238-8915.38298>. Disponível em: <https://seer.ufrgs.br/organon/article/view/38298>. Acesso em: 11 set. 2020.

HUBACK, A. P. *Efeitos de frequência nas representações mentais*. 2007. 318 p. Tese. Faculdade de Letras, Universidade Federal de Minas Gerais, 2007.

HUBACK, A. P. Plurais irregulares do português brasileiro: efeitos de frequência. *Revista da Abralin*, Curitiba, v. 9, n. 1, p. 11-40, 2010a. DOI: <https://doi.org/10.5380/rabl.v9i1.52337>. Disponível em: <https://revistas.ufpr.br/abralin/article/view/52337/32236>. Acesso em: 11 set. 2020.

HUBACK, A. P. Plurais em -ão do português brasileiro: efeitos de frequência. *Revista Linguística*, Rio de Janeiro, v. 6, n. 1, p. 9-26, 2010b. DOI: <https://doi.org/10.31513/linguistica.2010.v6n1a4436>

RIZZATO, É. *Interação do plural de -ão e do aumentativo -zão na formação de compostos no português brasileiro*. 2018. 94 f. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2018.

SCHWINDT, L. C. S. *et al.* A influência da variável escolaridade em fenômenos fonológicos variáveis: efeitos retroalimentadores da escrita. *Revista Virtual de Estudos da Linguagem – ReVEL*, [S.l.], v. 5, n. 9, p. 1-12, 2007. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/184784/000640837.pdf?sequence=1>. Acesso em: 11 set. 2020.

WETZELS, L. The Lexical Representation of Nasality in Brazilian Portuguese. *Probus*, [S.l.], v. 9, p. 203-232, 1997.

WETZELS, L. Comentários sobre a estrutura fonológica dos ditongos nasais no Português do Brasil. *Revista de Letras*, Fortaleza, v. 1, n. 22, p. 25-30, 2000. DOI: <https://doi.org/10.1515/prbs.1997.9.2.203>. Disponível em: <http://www.revistadeletras.ufc.br/rl22Art03.pdf>. Acesso em: 11 set. 2020.