

**Blending com nomes de animais em português brasileiro:
aspectos morfofonológicos e de frequência lexical**

**Blending with animal names in Brazilian Portuguese:
morphophonological and lexical frequency aspects**

Luiz Carlos Schwindt¹, Érika Roseli Pereira², Pedro Eugênio Gaggiola³

UFRGS-CNPq/Brasil, UFRGS/Brasil, UFRGS/Brasil

RESUMO

Este artigo trata do processo de formação de palavras denominado blending em dados escritos do português brasileiro em perspectiva descritivo-quantitativa. O objetivo do trabalho é discorrer sobre aspectos morfofonológicos e lexicais implicados na emergência desse fenômeno. A amostra é constituída por blends formados por nomes de animais, organizada a partir de diferentes fontes. As principais variáveis exploradas são *número de sílabas*, *tonicidade do blend* e *de suas bases e frequência lexical das bases*. Os resultados apontaram para predomínio de blends paroxítonos polissilábicos formados por bases paroxítonas dissilábicas ou trissilábicas. Ainda, a tonicidade do blend mostrou-se categoricamente equivalente à tonicidade da segunda base que o forma. Quanto à frequência, tomando-se como referência o Corpus Brasileiro, as bases identificadas como mais frequentes apresentaram menos apagamentos segmentais se comparadas às bases menos frequentes.

PALAVRAS-CHAVE:

Blend. Cruzamento vocabular. Formação de palavras. Morfologia. Morfofonologia.

ABSTRACT

This article addresses the word formation process known as blending in written data of Brazilian Portuguese from a descriptive-quantitative perspective. The objective of the study is to discuss the morphophonological and lexical aspects involved in the emergence of this phenomenon. Our data sample consists of blends formed from animal names, organized using different sources. The main variables explored are *number of syllables*, *blend stress* and *the stress of its bases*, and *lexical frequency of the bases*. The results show a predominance of polysyllabic blends with penultimate stress formed by disyllabic or trisyllabic bases with penultimate stress. Moreover, the stress of the blend was found to be categorically equivalent to the stress of its second base. As far as frequency is concerned, using Corpus Brasileiro as a reference, the bases identified as more frequent showed fewer segmental deletions when compared to less frequent bases.

KEYWORDS:

Blend. Vocabulary crossing. Word formation. Morphology. Morphophonology.

Recebido em: 02/08/2024

Aceito em: 10/12/2024

¹ schwindt@ufrgs.br | ORCID: <https://orcid.org/0000-0003-0533-589X>

² erikaroseli.pereira@gmail.com | ORCID: <https://orcid.org/0000-0002-7835-0102>

³ pedroe.gaggiola@gmail.com | ORCID: <https://orcid.org/0000-0002-0123-6205>

1. Introdução

Blending⁴ — fenômeno também conhecido como cruzamento vocabular (Basilio, 2003, 2005; Sandmann, 1988, 1991), palavra-valise (Alves, 1990), *portmanteau* (Araújo, 2000; Piñeros, 2000, 2004) e mistura (Sandalo, 2005) — é um processo de formação de palavras que consiste na junção de dois vocábulos já existentes, por exemplo, *chafé* (chá + café) e *apartamento* (apertado + apartamento). Apresenta características de processo produtivo no português Brasileiro (PB) e está intrinsecamente ligado à criatividade linguística, sendo geralmente encontrado em contextos informais do uso linguístico, como em conversas do dia a dia, redes sociais ou programas humorísticos.

Neste trabalho, que se soma a análises do fenômeno em diferentes perspectivas (Marangoni Jr., 2021; Minussi & Villalva, 2020; Villalva & Minussi, 2022; entre outros), queremos responder sobre quais são os motivadores envolvidos nas formações por blending quando estão em jogo nomes de animais. Nosso objetivo é descrever quantitativamente aspectos morfofonológicos e de frequência lexical envolvidos no processo fazendo uso de uma amostra original de dados escritos constituída por blends com essas propriedades. A análise que propomos se justifica pela demanda descritiva nesse segmento específico de dados.

Os trabalhos resenhados na seção 2 alimentam diretamente as hipóteses em investigação neste estudo. Hipotetizamos que (i) os blends da amostra apresentam a mesma posição acentual da segunda palavra base; (ii) os blends assumem o número de sílabas da base clipada⁵; (iii) os blends do tipo fonológico⁶ exibem, no ponto de encontro das bases⁷, o padrão *onset-rima*, enquanto os blends do tipo morfológico exibem o padrão *entre sílabas*; (iv) a base mais extensa do blend equivale ao clip e (v) as bases com maior frequência lexical correspondem às bases mais apagadas nos blends. Essa última hipótese fundamenta-se na premissa de que há relação diretamente proporcional entre frequência de uso e suscetibilidade à redução fônica, na linha do que sugere Bybee (2001) para diferentes fenômenos linguísticos.

Para testarmos as hipóteses elencadas, controlamos as seguintes variáveis, detalhadas na seção 3: (i) tonicidade dos blends e das bases, (ii) número de sílabas dos blends e das bases, (iii) estrutura do blend, (iv) tipo do blend, (v) estrutura silábica do encontro das bases, (vi) quantidade

⁴ Neste artigo, por critério de objetividade terminológica, e porque são expressões bem assimiladas na área, decidimos adotar os termos do inglês *blending*, para nomear o processo de formação de palavras, e *blend*, para nomear o produto final desse processo. Com base nessa decisão, dispensamos qualquer destaque na grafia dessas palavras.

⁵ Com base em Minussi & Villalva (2020).

⁶ Com base em Minussi & Nóbrega (2014).

⁷ Com base em Gonçalves (2003).

de segmentos fônicos apagados das bases e (vii) frequência lexical das bases.

O artigo está organizado como segue. Na seção 2, que segue esta introdução, contextualizamos o fenômeno frente aos nossos objetivos de modo não exaustivo a partir de alguns trabalhos referenciais sobre o tema. Na seção 3, tratamos da metodologia utilizada nesta pesquisa, descrevendo nossa coleta de dados, bem como detalhando as variáveis investigadas na análise. Na seção 4, apresentamos uma discussão acerca dos resultados obtidos por meio da descrição da distribuição de nossos dados. Na seção 5, por fim, sintetizamos as conclusões e limitações do trabalho e sinalizamos para perspectivas de continuidade da investigação.

2. Características morfofonológicas do blending e o papel da frequência lexical

Processos não concatenativos, como o blending, foram considerados por algum tempo como processos marginais ou irregulares em estudos sobre formação de palavras. Entretanto, o interesse pelo fenômeno vem se redimensionando nas últimas décadas, alimentando análises que assumem o desafio de descrever e formalizar o fenômeno, como as de Gonçalves (2003), Marangoni Jr. (2021), Minussi & Villalva (2020), entre outras. Na presente seção, realizamos um apanhado teórico a respeito de alguns aspectos morfofonológicos do blending presentes na literatura, bem como levantamos questões a respeito do papel da frequência das bases envolvidas nesse processo. Ao final, relacionamos mais diretamente a literatura visitada com os objetivos específicos deste estudo.

Gonçalves (2003) analisa blends do PB quanto ao tipo de encontro das bases. O autor observa que, geralmente, as palavras-base apresentam dois tipos de encaixe na mistura: no interior das sílabas, entre um onset da primeira base e uma rima da segunda base (ex. *maravilh/ástico* < *maravilhoso* + *fantástico*), ou entre duas sílabas, sem junção de elementos advindos de sílabas diferentes (ex. *brasi/nhol* < *brasileiro* + *espanhol*). Segundo Gonçalves & Assunção (2009, p. 68), todavia, há casos em que as duas palavras-base estão tão mescladas que é difícil definir quais segmentos pertencem a cada base individualmente, caracterizando-se tais segmentos, em sua interpretação, como compartilhados (ex. *comemorango* < *comemorando* + *morango*).

Outra abordagem do fenômeno é a de Minussi & Nóbrega (2014) e Nóbrega & Minussi (2015), que diferenciam blends levando em conta a presença ou a ausência de sobreposição fônica das bases. Os autores separam blends morfológicos, aqueles contendo apenas apagamento de segmentos de uma ou de ambas as bases (ex. *livemício* < *live* + *emício*), de blends que rotulam

como fonológicos, aqueles que apresentam, além de apagamentos, alguma similaridade fônica entre seus constituintes (ex. roubodízio < roubø + #odízio).

Minussi & Villalva (2020) agrupam blends quanto à estrutura mórfica dos seus constituintes, usando o rótulo *palavra* para um radical, um tema ou uma palavra completa presente no blend, e o termo *clip* para formas que não possuem, em tese, estatuto morfológico, representando um corte supostamente aleatório na palavra. Os autores dividem os blends em três grupos: (i) *clip*–palavra (CP) (ex. emprestado < emprestado + dado); (ii) palavra–*clip* (PC) (ex. paitrocínio < pai + patrocínio); e (iii) *clip*–*clip* (CC)⁸ (ex. abreijo < abraço + beijo). Em Villalva & Minussi (2022), analisaram quantitativamente 184 blends do *Portuguese Blend Corpus* e atestaram 39% de itens com padrão CP (ex. lambeijo < lambida + beijo), 37% com padrão CC (ex. galigumes < galinha + legumes) e 24% com padrão PC (ex. noivorado < noivo + namorado).

Além disso, Minussi & Villalva (2020) diferenciam blends cujas bases exibem relação semântica de modificação (ex. *morangoxinha*, coxinha do tipo morango, ou *capifruta*, caipirinha de fruta) de blends com bases relacionadas por coordenação, nos quais não cabe atribuir a um de seus constituintes o status de núcleo semântico (ex. *fabulástico*, algo igualmente fabuloso e fantástico). Nessa proposta, os blends que integram nossa amostra classificam-se, em sua totalidade, como blends coordenados.

Marangoni Jr. (2021) analisou um corpus de blends composto por 411 dados coletados de diversas fontes. Em relação ao número de sílabas, constatou 41% de itens em que o produto do processo apresentava o mesmo número de sílabas da segunda base formadora do blend (ex. co.po.nhei.ro < co.po + com.pa.nhei.ro) e 59% de itens em que o produto apresentava o mesmo número de sílabas da primeira base (ex. ca.ri.ú.cho < ca.ri.o.ca + ga.ú.cho) ou um número de sílabas diferente das duas bases (ex. a.ta.ca.re.jo < a.ta.ca.do + va.re.jo)⁹. Com relação ao acento, a maior parte dos dados da amostra contribuiu para a ideia de que é a segunda base que exerce papel sobre a atribuição de acento do blend (ex. toboágua < tobogã + água). Considerada a distinção entre blends fonológicos e morfológicos, o autor identifica em sua análise 235 casos do primeiro e 159 casos do segundo.

Medir aspectos de frequência de blends não é tarefa trivial. Grande parte desses vocábulos possui vida efêmera, com algumas exceções, como *portunhol*, que pode já estar lexicalizado no PB,

⁸ Villalva & Minussi, em comunicações orais recentes, vêm sugerindo o acréscimo de um tipo de blend à classificação proposta em Minussi & Villalva (2020), palavra-palavra, o qual incorporamos a esta análise. Nesse tipo, as duas bases estão visíveis e se registra sobreposição fônica (ex. escaravelhota < escaravelho + velhota).

⁹ Por critério de clareza e simplicidade de exposição, e por não ter implicações sobre os argumentos, empregamos pontos para indicar limites de sílabas sobre formas ortográficas neste texto. O critério para a separação é fonético, contudo, e se orienta pela divisão que consideramos mais comum dessas palavras em PB.

isto é, é possível que não seja mais identificada transparentemente sua estrutura interna, de modo a descaracterizar-se o processo de formação por *blending* propriamente dito. Entretanto, com o ambiente propício para o registro escrito de neologismos em redes sociais, vem se ampliando o uso e a percepção do uso dessas criações vocabulares. Ainda assim, registrar a frequência de ocorrência de *blends* apresenta desafios. Villalva & Minussi (2022), por exemplo, utilizaram o *Corpus Now* para coletar a frequência dos *blends* de sua amostra. O estudo registra que grande parte dos *blends* avaliados exibia frequência de *tokens* igual a zero ou 1 (ex. *anãofabeto*, 0; *caligrafeia*, 1), enquanto uma pequena parcela exibia um número razoável de ocorrências (ex. *cantautor*, 752). Isso sugere que uma das alternativas para compreendermos adequadamente o papel da frequência nessas formações é olhar com cuidado para a frequência de seus constituintes, metodologia adotada em outros estudos quantitativos sobre *blending* (Kelly, 1998; Villalva & Minussi, 2022). Essa decisão metodológica fundamenta-se na premissa de que a frequência das bases que formam o *blend* pode influenciar seu reconhecimento e formação. *Blends* criados a partir de constituintes mais frequentes tendem a ser mais facilmente recuperados, como demonstrado na literatura sobre o fenômeno (Lehrer, 1996; Juhasz, Johnson & Brewer, 2017).

Kelly (1998, p. 580) analisa 320 formações do inglês com objetivo de investigar a ordem em que as bases aparecem no *blending*. O autor observa que a escolha depende de critérios como frequência, estrutura prosódica e alguns aspectos semânticos das palavras-base. Para reforçar essa ideia, Kelly (1998) traz alguns exemplos de *blends* do inglês, como *donkephant*, que poderia ter vencido a forma *eledonk* porque *donkey* (burro) tem menos sílabas que *elephant* (elefante), ou como *smog*, que poderia ter vantagem sobre *foke*, pelo fato de que *smoke* (fumaça) apresenta maior frequência do que *fog* (névoa). Seus resultados sugerem, de modo geral, que a primeira base da mistura tende a ser mais frequente e menor em extensão silábica. O estudo de Kelly (1998) é coerente com as principais ideias da Fonologia de Uso (Bybee, 2001, 2010), modelo segundo o qual a frequência de uma palavra determina, em grande medida, sua representação linguística e os processos aos quais ela se submete.

3. Metodologia

3.1 Coleta de dados

Com o objetivo de quantificar propriedades morfológicas, fonológicas e de frequência do

blending em PB, propomos a descrição de uma amostra de 201 blends coordenados formados a partir de nomes de animais. A amostra encontra-se integralmente disponível no Quadro 01 abaixo.

Quadro 01 – Amostra de blends escritos formados a partir de nomes de animais

Aguielho, bacaelho, porquelho, rinocoelho, tamanduelho, passarneiro, cavaleão, escorleão, galeão, botucano, macacano, orangocano, patucano, tatucano, tubacano, golfira, rinocerra, esquilobo, gorilobo, macacobra, peicobra, porcobra, rinocerobra, urucobra, maripeixe, gatavalo, gatovalo, pandavalo, urubaca, formicão, porcão, zebrarata, jiboi, pomboi, uruboi, carangato, minhogato, morcegado, pulgado, vacato, papachorro, patorro, tartachorro, cangurinha, tubarana, cachorrato, cigarrato, borbolecao, emacaco, giracaco, tourave, joalinha, vacalinha, centopato, chimpampato, hipopato, sapato, gansubarão, jacoruja, macacoruja, porcoruja, ursoruja, castouro, tartatouro, cavarinho, vacarinho, jabotalo, vacalo, borbofante, formigante, girafante, hipofante, tartafante, tubafante, perniguica, peruguica, avestigre, quatigre, tucastor, caracoleta, carneleta, cobroleta, crocoleta, esquileta, hipoleta, macacoleta, marmoleta, minholeta, porcoleta, ursoleta, vacaleta, cangurumiga, cobrormiga, carnurso, minhocurso, raturso, sapurso, uruburso, hipopotosa, saposa, elefantema, jacarema, camelha, jacabelha, pombelha, urubelha, bodonça, borbolonça, camundonça, grilonça, jaconça, macaconça, polvonça, rinoceronça, tamanduonga, zebronça, garcaruga, hiporuga, tucaruga, cangaré, gambaré, saporé, tubaré, cabragaio, cangugaio, tubagaio, jacarágua, cãoleia, cobraleia, hipopotaleia, jacaleia, leopavão, sapavão, burroca, elefoca, girafoca, girafossauro, tartarejo, chimpanzebra, formiguru, gambaru, pinguru, sapuru, giramelo, macacamelo, moscamelo, vacamelo, dromestruz, porcodilo, cabrubu, perurubu, cangurujo, cobronha, morcegonha, rinocegonha, cabrinhoca, papanhoca, tartanhoca, araranta, tucananta, avestruta, crocogrilo, macacol, cacholvo, porvo, leopanda, gatorinha, colibrito, pavanduá, formiganso, tubaranso, galorila, lebelula, camelontra, leontra, dinossirafa, peirafa, gorilatixa, tatutixa, porcopião, elefanhoto, patanhoto, tamanhoto, cobroceronte, esquiceronte, morceronte, onceronte, burreco, barapótamo, camelótamo, esquilótamo, patopótamo, peripótamo, raposótamo, mosquito, abelhama, ovelhama, peliena, vacalume, sirilampo, camundoninha.

Fonte: Elaboração própria

A análise de frequência foi baseada na frequência lexical de cada base do blend, extraída

do Corpus Brasileiro (CBras)¹⁰. O CBras conta com dados provenientes de fontes de língua escrita e falada e apresenta cerca de um bilhão de *tokens* e quase seis milhões de *types*, caracterizando-se como um robusto *corpus* de referência. As fontes utilizadas na coleta de dados são as seguintes: (i) figurinhas do chiclete Ping Pong, destinadas ao álbum *Floresta Maluca*, comercializado na década de 1970; (ii) figurinhas da coleção *Ploc o Chicle Divertoso*, comercializadas na década de 1990; (iii) livro infantil *Misturichos*, escrito por Beatriz Carvalho & Renata Bueno, publicado em 2012; (iv) livro infantil *Ajidar, o dragão da terra*, escrito por Marjane Satrapi, publicado em 2007; (v) blog da professora Liane Castro de Araujo, que descreve a brincadeira *Bichionário*; (vi) música infantil *Misturando os Bichos*, com composição de Bruno Cardoso Santoro. A seleção das fontes utilizadas iniciou-se restrita às figurinhas dos chicletes Ping Pong e Ploc. Em uma segunda instância da coleta de dados, para que a amostra alcançasse um número mais consistente de blends, incluíram-se dados advindos das fontes descritas em (iii-vi). Todas as fontes utilizadas continham pistas visuais a respeito dos constituintes formadores dos blends, como se observa na Figura 1.

Figura 1 – Figurinhas *urubelha*, *saparé*, *tamanhoto* e *jacoruja*



Fonte: Coleção de Figurinhas Chiclete Ploc e Chiclete Ping Pong

A escolha de dados dessa natureza, que se situam pragmaticamente em contexto de ludicidade, é justificada pela ideia de que tudo o que se produz numa língua natural em termos de formação de palavras é reflexo, na sua medida, da competência lexical dos falantes.¹¹

¹⁰ <https://www.linguateca.pt/aceso/corpus.php?corpus=CBRAS>

¹¹ Competência lexical é entendida aqui como conhecimento internalizado dos usuários a respeito da estrutura interna dos vocábulos de sua língua e dos processos/regras que manipulam morfemas, independentemente da natureza do significado veiculado por essas estruturas.

3.2 Variáveis e distribuição dos dados

Após a coleta de dados, listamos e codificamos a amostra conforme as variáveis que seguem.

Tabela 1 – Variáveis analisadas

Variáveis	Exemplos	
(i) tonicidade do blend e das bases	a. proparoxítono b. paroxítono c. oxítono	a. barapótamo b. jacaleia c. cabrubu
(ii) número de sílabas do blend e das bases	a. Uma sílaba b. Duas sílabas c. Três sílabas d. Quatro sílabas e. Cinco sílabas f. Seis sílabas	a. rã b. gri.lo c. já.bu.ti d. per.ni.lon.go e. o.ran.go.tan.go f. hi.po.po.ta.lei.a
(iii) estrutura do blend	a. CP b. OC c. CC d. PP	a. tartatouro b. patopótamo c. dinossirafa d. morcegonha
(iv) tipo de blend ¹²	a. fonológico b. morfológico	a. maripeixe: mariposa + peixe b. rinocoelho: rinoceronte + coelho
(v) estrutura silábica do encontro das bases	a. onset–rima b. entre sílabas	a. morceg/ato: morcego + gato b. orango/cano: orangotango + tucano
(vi) quantidade de segmentos fônicos apagados das bases	De 0 a 7 segmentos	/ipopotamo/ e /borboleta/ base um: 7 segmentos apagados base dois: 4 segmentos apagados
(vii) frequência lexical das bases	Base menos frequente: 30 Base mais frequente: 50,849	doninha – 30 coelho – 50,849

Considerando essas variáveis, descrevemos e analisamos a distribuição de nossos dados fazendo uso da Plataforma R (R Core Team, 2024), por meio da interface *RStudio* (RStudio Team, 2024). Os gráficos foram gerados em *ggplot2* (Wickham, 2009).

¹² Diferentemente do que propõem Minussi & Nóbrega (2014) e Nóbrega & Minussi (2015), em nossa amostra não se atestaram blends *semânticos*. Em relação a essa variável, optamos, ainda, por olhar apenas para sobreposição fônica no ponto de encontro das bases, seguindo Marangoni Jr. (2021, p. 166).

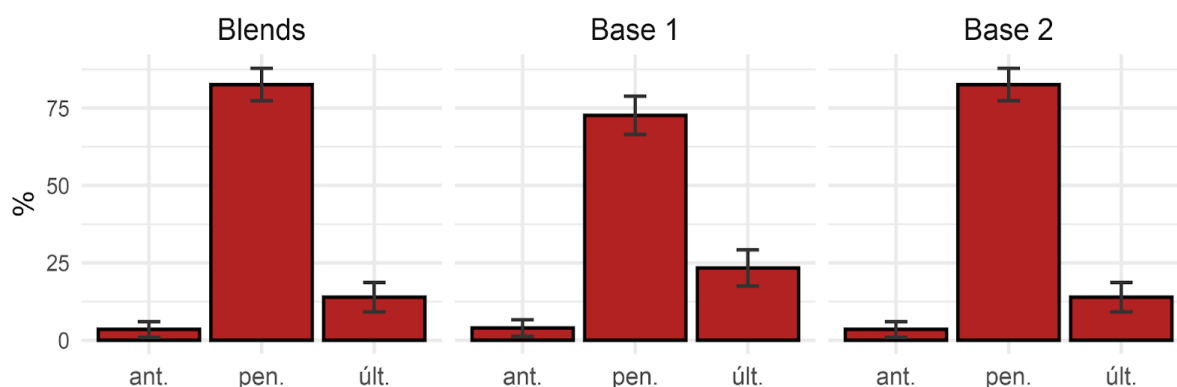
4. Resultados e discussão

Nesta seção apresentamos os principais resultados de nossa análise quantitativa, acompanhados de considerações de natureza predominantemente descritiva. Para melhor avaliarmos as diferenças na distribuição dos dados, os gráficos de proporção subsequentes encontram-se complementados por barras de erro que indicam os intervalos de confiança (IC's) considerando $\alpha = 5\%$, valor assumido também nos testes estatísticos aplicados e reportados a seguir.

4.1 Acento

Com relação ao acento, a maior parte dos blends e de palavras-base analisados é paroxítona, enquanto estruturas proparoxítonas são minoritárias, como se vê na Figura 2. Esses resultados seguem as tendências gerais do acento em português (Camara Jr., 1970).

Figura 2 – Tonicidade das bases e dos blends (n = 603)



Fonte: elaborado pelos autores

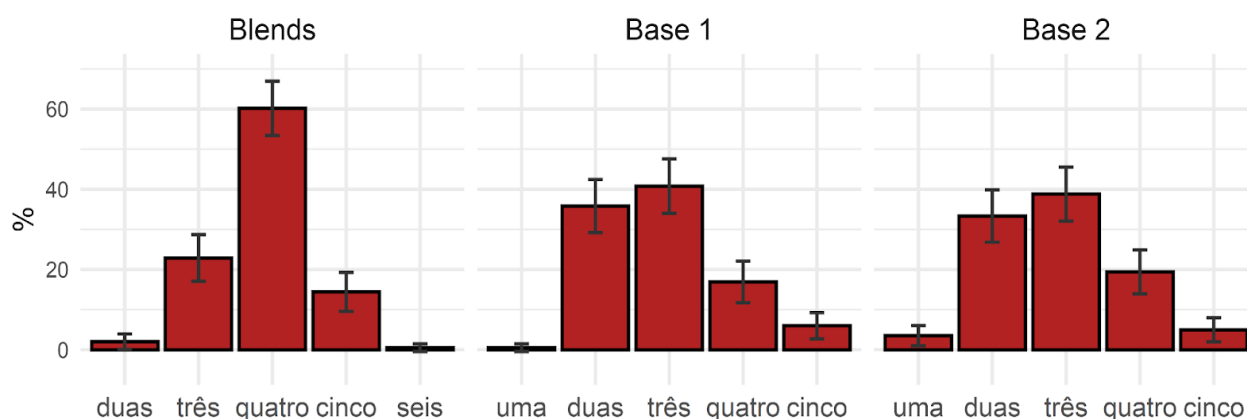
Além disso, em todos os blends avaliados atestamos coincidência da tonicidade do blend com a tonicidade da sua segunda base (ex. ra.po.só.ta.mo < ra.po.sa + hi.po.pó.ta.mo), evidenciada na comparação da posição do acento do blend e da base dois, resultados similares aos de Marangoni Jr. (2021). A influência determinística do padrão acentual da segunda base no padrão acentual dos blends pode ser relacionada ao fato de que o acento em português recai obrigatoriamente sob uma das três últimas sílabas do vocábulo. Em decorrência disso, não seria bem formada uma estrutura que levasse consigo o acento da primeira base em casos como *cobraleia*, por exemplo (ex. *có.bra.lei.a).

Observamos, ainda, que os dados da primeira e da segunda bases distribuem-se de maneira semelhante quanto à posição do acento, conforme a Figura 2. Mesmo assim, o teste de qui-quadrado mostrou haver diferença significativa entre as proporções ($\chi^2= 6,16(2)$, $p < 0,05$). O exame dos valores de χ^2 por pares de células indica que bases oxítonas são as principais responsáveis por essa diferença ($\chi^2_{\text{últ.}}= 2,40$; $\chi^2_{\text{pen.}}= 0,64$; $\chi^2_{\text{ant.}}= 0,03$).

4.2 Número de sílabas

Com relação ao número de sílabas das bases e do blend, na amostra predominam blends polissilábicos de quatro sílabas, caracterizando-se a maior parte das bases como dissilábicas ou trissilábicas (ex. bu.rro.ca < bu.rro + fo.ca; ba.ca.e.lho < ba.ca.lhau + co.e.lho). A sobreposição de IC's para bases com duas ou três sílabas indica não haver diferença importante na distribuição dos dados quanto ao fato de as bases serem dissilábicas ou trissilábicas. Ainda, não há registro de blends monossilábicos na amostra e, mesmo entre as bases, monossílabos são raros. As bases um e dois apresentam no máximo cinco sílabas e só há um blend na amostra com seis sílabas, *hi.po.po.ta.lei.a* (hipopótamo + baleia). Os resultados podem ser conferidos na Figura 3.

Figura 3 – Número de sílabas das bases e dos blends (n = 603)



Fonte: elaborado pelos autores

Em uma comparação descritiva entre o número de sílabas do blend e o número de sílabas de suas bases, observamos que 51,2% dos blends apresentam um número de sílabas diferente (maior ou menor) do que suas palavras-base (ex. ra.tur.so < ra.to + ur.so), 26,4% assumem o número de sílabas da sua segunda base (ex. va.ca.lu.me < va.ca + va.ga.lu.me), 17,4% assumem o número de sílabas da sua primeira base (ex. u.ru.boi < u.ru.bu + boi) e apenas 5% dos blends

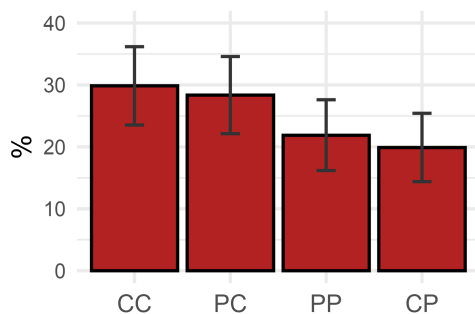
apresentam o mesmo número de sílabas de ambas as bases envolvidas no processo (ex. por.vo < por.co + pol.vo). Esse achado contraria a hipótese de que o blend tende a assumir o mesmo número de sílabas que sua segunda base formadora (Bat-El, 2006) e difere de demais achados presentes na literatura do fenômeno a respeito da comparação entre o número de sílabas de blends e suas bases (Marangoni Jr., 2021).

Além disso, comparando a extensão silábica das bases dos blends, verificamos que em 37,3% a primeira é a mais extensa (ex. ri.no.ce.go.nha < ri.no.ce.ron.te + ce.go.nha), em 36,8%, a segunda (ex. ca.me.ló.ta.mo < ca.me.lo + hi.po.pó.ta.mo) e em 25,9% as duas são de igual tamanho (ex. cas.tou.ro < cas.tor + tou.ro). Avaliamos também a média de número de sílabas de cada base, que corresponde a 2,92 e 2,89 sílabas para a primeira e a segunda, respectivamente. A mediana do número de sílabas correspondeu a três para ambas as bases, diferente do registrado por Kelly (1998) para dados do inglês, que constatou médias de 2,2 e 2,7 para a primeira e a segunda, respectivamente. As medidas de tendência central identificadas em nosso trabalho não permitem que diferenciemos a primeira e a segunda base formadora de blends de maneira importante quanto à extensão silábica.

4.3 Estrutura do blend

Analisando a distribuição dos dados em função da estrutura do blend, não observamos predominância significativa de nenhum dos padrões de estrutura avaliados, resultado deduzido a partir da sobreposição de IC's (conforme Figura 4) e corroborado por um teste de qui-quadrado ($\chi^2 = 5,66(3)$, $p > 0,1$). Em números absolutos, observamos que a maior parte da amostra é constituída de blends *clip-clip* (ex. rinocerobra < rinoceronte + cobra), seguidos de *palavra-clip* (saposa < sapo + raposa), *palavra-palavra* (tatuano < tatu + tuano) e, por fim, *clip-palavra* (urucobra < urubu + cobra). A predominância de blends CC, sem significância estatística em nossos resultados, iria ao encontro da hipótese de que blends coordenados são formados idealmente por dois *clips*, de acordo com a descrição de Villalva & Minussi (2022). Esse resultado, contudo, pode estar comprometido por limitações da amostra de que nos valem neste trabalho, seja pela natureza particular dos dados seja pelo volume relativamente restrito de itens.

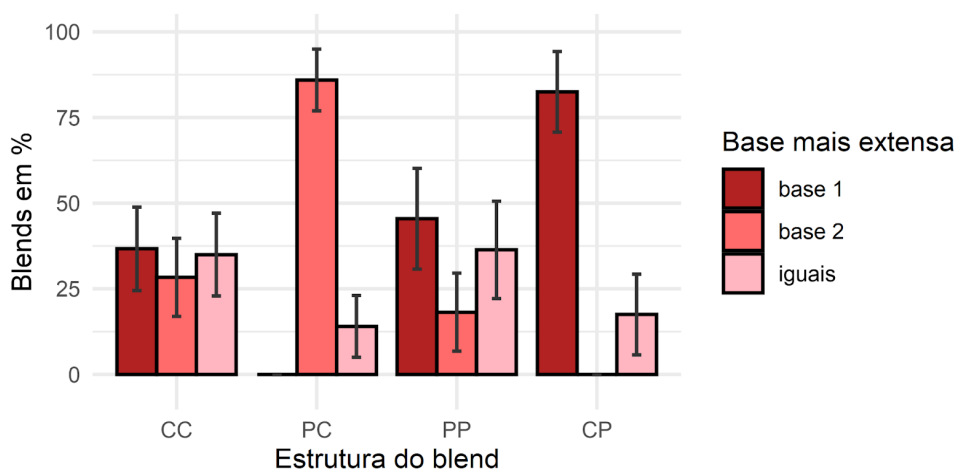
Figura 4 – Estrutura do blend (n = 201)



Fonte: elaborado pelos autores

Ao observarmos a relação entre a estrutura do blend e a base de maior extensão silábica da mistura, na Figura 5, notamos que blends PC apresentam, em sua maioria, a segunda base como a mais extensa (ex. *gambaru* < *gambá* + *canguru*) e blends CP apresentam majoritariamente a primeira base como a mais extensa (ex. *camundonça* < *camundongo* + *onça*), resultados evidenciados pela não sobreposição de barras de erro nesses grupos. Em ambas as classificações, portanto, a palavra-base de maior extensão corresponde à palavra clipada no blending. Outro achado interessante é que em blends PC e CP não se registram dados nos quais P se apresente como a base mais extensa. Esses achados parecem sustentar a ideia de que, na formação de um blend, é preferível preservar a menor das bases e apagar segmentos da base maior, assegurando pelo material fônico a maior inteligibilidade possível de seu significado. Em blends CC e PP, os dados distribuem-se de maneira mais uniforme quanto à base que figura como a mais extensa.

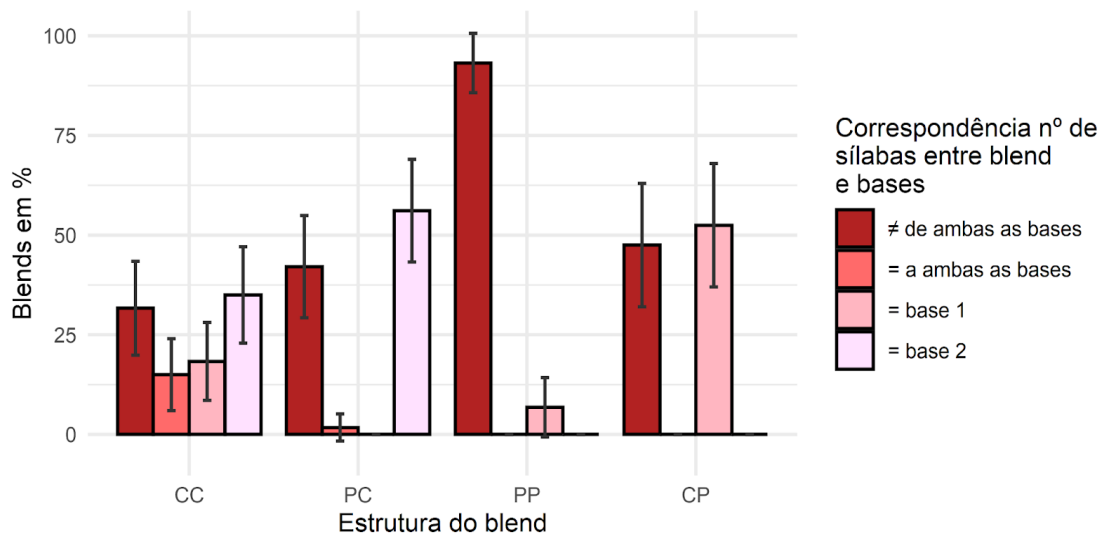
Figura 5 – Relação entre estrutura do blend e base de maior extensão silábica (n = 201)



Fonte: elaborado pelos autores

A Figura 6, a seguir, descreve a distribuição dos dados considerando-se o cruzamento entre estrutura do blend e correspondência de número de sílabas de blends e bases.

Figura 6 – Relação entre estrutura do blend e correspondência de nº de sílabas entre blend e bases (n = 201)



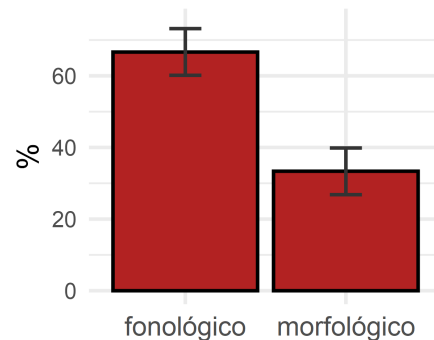
Fonte: elaborado pelos autores

Ao relacionarmos essas duas variáveis, constatamos que a maior parte dos blends PC apresenta extensão silábica igual à base dois ou diferente de ambas as bases (ex. pa.to.pó.ta.mo < pa.to + hi.po.pó.ta.mo; ur.so.ru.ja < ur.so + co.ru.ja). Já blends CP apresentam extensão silábica igual à base um ou diferente de ambas as bases (ex. for.mi.cão < for.mi.ga + cão; a.ves.ti.gre < a.ves.truz + ti.gre). Em outros termos, pode-se dizer que o blend tende a assumir o mesmo número de sílabas da base que corresponde ao *clip*, e não à palavra, nos casos de CP e PC. Esse resultado pode indicar que reproduzir o número de sílabas do constituinte mais apagado no blend é uma forma de facilitar a recuperação desse constituinte. Já em blends PP, ou seja, naqueles que apresentam as bases completas com sobreposição fônica, a maioria dos itens exibe um número de sílabas diferente do número de sílabas de suas bases (ta.tu.ca.no < ta.tu + tu.ca.no). Em blends CC, os dados distribuem-se de maneira mais uniforme. Resultados similares foram encontrados em Villalva & Minussi (2022).

4.4 Tipo de blend

Na Figura 7, a seguir, atestamos que a maior parte dos blends é do tipo fonológico (u.ru.bur.so < u.ru.bu + ur.so), ou seja, exibe algum compartilhamento fônico no encontro das bases ($\chi^2= 22,33(1)$, $p < 0,001$).

Figura 7 – Tipo de blend (n = 201)



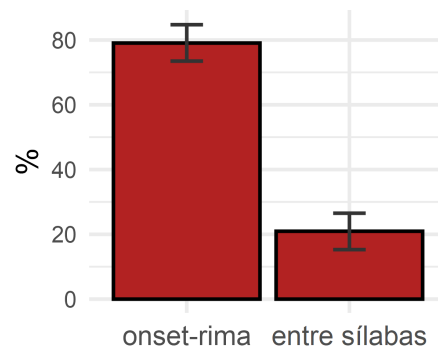
Fonte: elaborado pelos autores

Esse resultado vai ao encontro do achado de Marangoni Jr. (2021), uma vez que em sua amostra 57% dos blends eram do tipo fonológico. O achado também vai ao encontro da proposta de Gonçalves (2006, p. 178), na qual o autor assume que o compartilhamento de segmentos fônicos parece ser uma das principais motivações para criação de blends no PB. Esse quadro analítico nos conduz à ideia de que bases com semelhança fônica possuem maior chance de participar na formação de um blend, havendo contexto favorável para criação da nova palavra. Nesse sentido, a similaridade fônica entre as palavras-base parece apresentar papel importante no processo de blending (Gries, 2004), sendo essa uma característica que diferencia esse tipo de formação de palavras dos processos de composição ou afixação, por exemplo, nos quais não parece ser plausível assumir que a semelhança de segmentos fônicos possua papel relevante na combinação de radicais ou afixos.

4.5 Estrutura silábica do encontro das bases

A maioria dos blends apresenta no encontro das bases a estrutura silábica *onset-rima* (ex. tu.ca.n/an.ta < tu.ca.no + an.ta), conforme Figura 8 ($\chi^2= 68,10(1)$, $p < 0,001$). Já o padrão *entre sílabas* foi observado em pequena parte da amostra (ex. tu.ba./fan.te < tu.ba.rão + e.le.fan.te).

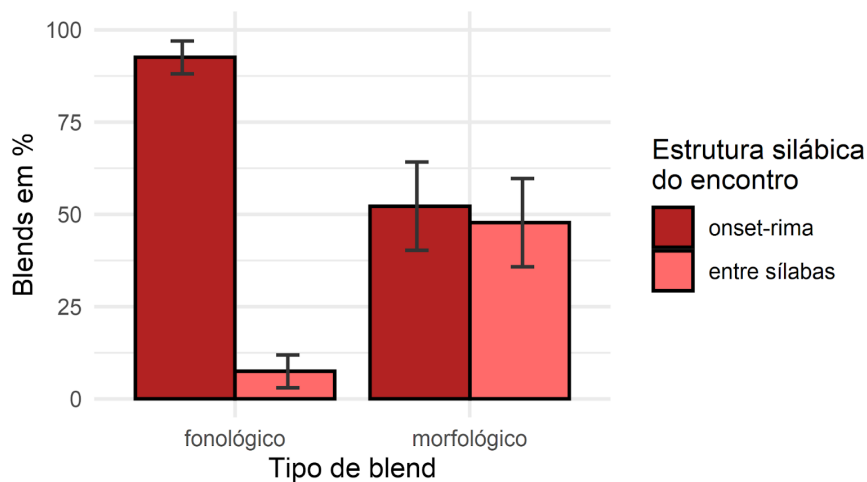
Figura 8 – Estrutura silábica do encontro das bases (n = 201)



Fonte: elaborado pelos autores

A Figura 9 mostra a relação entre o tipo de blend e o padrão silábico que resulta do encontro das bases implicadas nessas formações.

Figura 9 – Relação entre tipo de blend e estrutura silábica do encontro das bases (n = 201)



Fonte: elaborado pelos autores

Constatamos que a maior parte dos blends fonológicos apresenta o padrão *onset-rima* (ex. go.ri.l/o.bo < go.ri.la + lo.bo), com raros casos de blends que apresentam o padrão *entre sílabas*. Em blends morfológicos, por outro lado, não se observa distinção significativa entre as estruturas silábicas *onset-rima* e *entre sílabas*, achado validado por um teste de qui-quadrado ($\chi^2 = 0,13(1)$, $p > 0,7$) e pela sobreposição de barras de erro visualizadas na Figura 9. Esse achado sugere que há maior tendência de alinhamento entre constituintes morfológicos e fonológicos quando não há similaridade fonológica entre as bases — o caso de blends morfológicos (ex. bacaelho). Blends do

tipo fonológico, por outro lado, parecem privilegiar formações nas quais a sílaba presente no ponto de encontro entre as bases apresenta elementos de ambos os constituintes morfológicos que formam o blend (ex. ca.va.l/e.ão < ca.va.lo + le.ão).

4.6 Frequência lexical das bases

Dado o uso reduzido e contextualizado de blends, a análise de frequência lexical deste trabalho, em lugar de considerar a frequência das palavras resultantes do processo, orientou-se pela frequência de *tokens* de suas palavras-base, valendo-se dos índices fornecidos pelo CBraS, nossa amostra de referência. Objetivando verificar se palavras mais frequentes apareciam mais em primeira ou em segunda posição no blend, realizamos uma comparação entre as bases em cada formação, em termos de frequência lexical, e observamos que 56,2% dos blends apresentam a segunda base como a mais frequente (ex. pulgato < pulga, 0 + gato, 0,527)¹³, enquanto 43,7% possuem a primeira base como a mais frequente (ex. pavandua < pavão, 1,6 + tamanduá, 0,482). Ao contabilizarmos o número de ocorrências por milhão das palavras-base, atestamos uma média da frequência de *tokens* de 3,79 para as primeiras bases e de 5,99 para as segundas bases. Esse resultado se diferencia dos encontrados por Kelly (1998) para o inglês. O autor constatou uma média de 40,1 ocorrências por milhão para as primeiras bases e de 14,8 para as segundas bases. Esses resultados não permitem que façamos afirmações seguras a respeito do papel da frequência na determinação da posição dos constituintes que participam do blending em uma perspectiva que compara o inglês e o português, por exemplo.

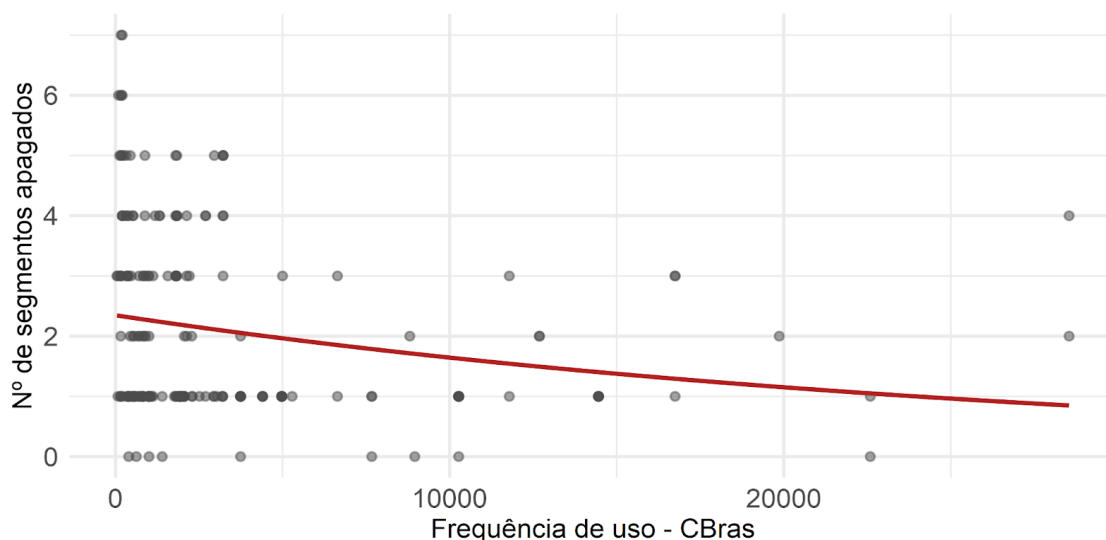
Inicialmente, tínhamos como hipótese que a relação entre apagamento de substância fônica e frequência da palavra base seria caracterizada de maneira diretamente proporcional: quanto maior a frequência da palavra-base, mais permissiva à perda de conteúdo fônico ela seria, numa leitura do que propõe Bybee (2001) para diferentes fenômenos fonético-fonológicos. Entretanto, nossos resultados apontam para uma conclusão contrária a respeito da relação entre apagamento de substância fônica e frequência. Observamos uma relação inversamente proporcional entre quantidade de segmentos apagados e frequência de *tokens*, quanto mais frequente é a base, menos apagamento ela apresenta. Esse resultado contraria nossa hipótese inicial, permitindo uma nova interpretação sobre o papel da variável, que sugere que é a ação conjunta de frequência e preservação de material fonético de uma das bases que assegura sucesso

¹³ As frequências reportadas nesse e no próximo exemplo de nossa amostra correspondem a valores de frequência por milhão.

na interpretação semântica adequada dos blends.

Essa relação entre número de segmentos fônicos apagados e frequência de *tokens* das bases está retratada nas Figuras 10 e 11, que fazem referência, nesta ordem, às primeiras e às segundas bases dos blends. Realizamos uma regressão de Poisson para avaliar a tendência observada que apontou para a significância estatística da relação em direção oposta entre número de segmentos fônicos apagados e frequência tanto para a base um ($\beta = -3.562^{-05}$, $SE = 1.217^{-05}$, $p < 0,001$) quanto para a base dois ($\beta = -3.262^{-05}$, $SE = 9.400^{-06}$, $p < 0,001$)¹⁴. Observamos, em síntese, que, conforme aumenta a frequência de *tokens* das bases, o número de segmentos apagados pelo blending tende a ser menor.

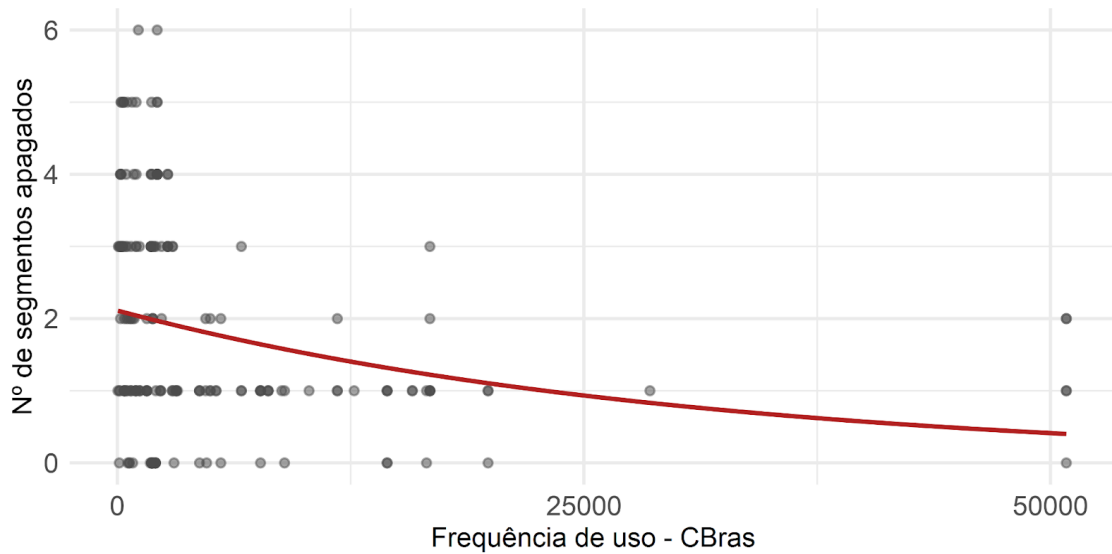
Figura 10 – Relação entre quantidade de segmentos fônicos apagados e frequência da base 1



Fonte: elaborado pelos autores

¹⁴ Realizamos também um teste de razão de verossimilhança comparando nossa regressão de Poisson com um modelo binomial negativo para verificar se haveria excesso de dispersão em nossos dados (Winter, 2020). O teste não se mostrou estatisticamente significativo na comparação entre os modelos dos dados da primeira base ($\chi^2 = 0.66(1)$, $p > 0,2$) e tampouco da segunda base ($\chi^2 = 2.30(1)$, $p > 0,06$). Assim, a relação observada nas Figuras 9 e 10 é mais bem entendida com o auxílio da regressão de Poisson relatada, uma vez que o modelo não apresenta sobredispersão.

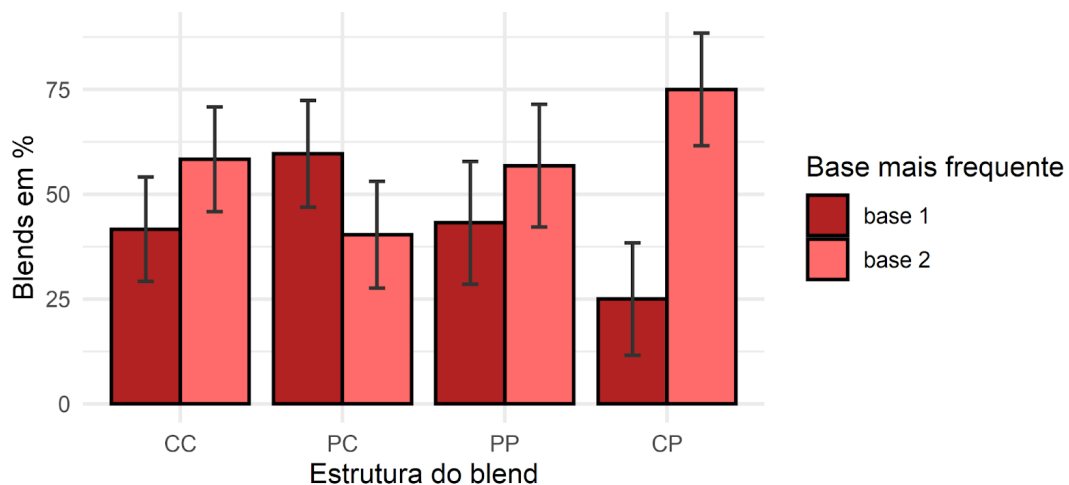
Figura 11 – Relação entre quantidade de segmentos fônicos apagados e frequência da base 2



Fonte: elaborado pelos autores

Outro resultado na direção da correlação aqui discutida está apresentado na Figura 12, que compara a distribuição dos dados em função da estrutura do blend e de sua base mais frequente.

Figura 12 – Relação entre estrutura do blend e base mais frequente do blend (n = 201)



Fonte: elaborado pelos autores

Observamos nos testes de qui-quadrado que blends com estrutura CC ($\chi^2 = 1,66(1)$, $p > 0,1$), PC ($\chi^2 = 2,12(1)$, $p > 0,1$) ou PP ($\chi^2 = 0,81(1)$, $p > 0,3$) não se diferenciam de maneira significativa, no interior de cada tipo, quanto à frequência das duas bases que os compõem. Já em blends CP, a

base que corresponde a P costuma ser a mais frequente ($\chi^2= 10(1)$, $p < 0,01$). Nesse caso em particular, podemos arriscar a interpretação de que estaria em jogo a preservação da integridade da base mais frequente, a de informação mais facilmente recuperável em dado sentido.

Assim, os principais resultados deste estudo parecem sugerir algumas generalizações acerca do presente fenômeno analisado. No que diz respeito ao acento, o padrão acentual paroxítono com preservação da posição acentual da palavra-base mais à direita é o preferido (ex. ja.ca.lei.a < jacaré + ba.lei.a), confirmando resultados prévios na literatura do fenômeno (Marangoni Jr., 2021). Já em relação ao número de sílabas, os blends da amostra são, em sua maioria, polissílabos formados a partir de bases dissilábicas ou trissilábicas (ex. gri.lon.ça < gri.lo + on.ça; for.mi.gu.ru < for.mi.ga + can.gu.ru). No que tange à estrutura do blend, não se observaram diferenças significativas na distribuição dos padrões de estrutura, ainda que, em números absolutos, o padrão *clip–clip* predomine (ex. crocoleta < crocodilo + borboleta), resultado que poderia corroborar o achado de Villalva & Minussi (2022).

A maior parte dos blends compartilha material fônico no encontro das bases (isto é, pode ser classificada como blend fonológico), embora haja uma parcela razoável de blends sem semelhança fônica entre suas palavras-base. No encontro das bases, a maioria dos blends apresenta contato intrassilábico, ou seja, o onset de uma sílaba da primeira base se une à rima de uma sílaba da segunda base (ex. jac/onça < jacaré + onça). Esses resultados vão ao encontro dos achados de Gonçalves (2003) e corroboram a ideia de que a similaridade fonológica entre as bases tem importante papel na seleção dos constituintes para que se realize blending (Gries, 2004).

Na maior parte dos blends CP, a primeira base é a mais extensa em termos de número de sílabas (ex. uruboi < urubu + boi). Em grande parte de blends PC, por outro lado, é a segunda base que é mais extensa (vacalhinha < vaca + galinha). Esse resultado vai ao encontro da ideia de que se prefere apagar material fônico da base de maior extensão silábica no blending. Além disso, concluímos que blends CP assumem majoritariamente o número de sílabas do *clip* ou um número de sílabas diferente de ambas as bases (ex. rinocoelho < rinoceronte + coelho). Esse resultado também é atestado em misturas do tipo PC (porcopião < porco + escorpião). Assim, os blends parecem tender a exibir o número de sílabas da base clipada (mas não da palavra), resultado similar àquele encontrado na amostra de Villalva & Minussi (2022). Essa tendência pode auxiliar na recuperação da base clipada uma vez que sua extensão silábica está preservada no blend.

Verificamos também que quanto maior é a frequência da palavra-base, menor é o número de segmentos apagados no resultado do blending. Opondo à suspeita de correspondência entre maior frequência e conseqüente maior chance de apagamento de segmentos fônicos, esse

resultado sugere que é a associação de frequência à preservação de substância fônica que favorece a interpretação semântica adequada da palavra formada. Nessa mesma direção, no padrão CP, a base mais frequente em termos de *tokens* é P, que, por sua vez, é também a base com menor número de apagamentos se comparada a C. Estudos futuros que disponham de maior número de dados podem chegar ao mesmo resultado para o padrão PC, como sugerido em termos descritivos pelos resultados aqui reportados.

5. Conclusão

Neste trabalho apresentamos um estudo descritivo que tomou como base uma amostra de 201 blends coordenados formados a partir de nomes de animais. Investigamos as seguintes variáveis de natureza morfofonológica e lexical: (i) tonicidade, (ii) número de sílabas, (iii) estrutura morfológica, (iv) tipo do blend, (v) estrutura silábica no encontro das bases, (vi) quantidade de segmentos fônico apagados e (vii) frequência lexical. Os resultados obtidos no estudo nos permitem algumas generalizações acerca do conjunto de dados analisado. Atestamos que os blends avaliados possuem o mesmo padrão acentual de sua segunda base e um número de sílabas diferente de qualquer uma das bases. Observamos, ainda, relação entre blends fonológicos e o tipo *onset-rima* e constatamos não haver diferenças importantes na distribuição dos dados no âmbito de blends morfológicos quanto à estrutura silábica do ponto de encontro das bases. Concluimos que a base mais extensa do blend tende a corresponder à base clipada, sendo também essa base a que cede sua extensão silábica ao produto do blending. Por fim, verificamos que há relação inversamente proporcional entre frequência lexical e número de segmentos fônicos apagados no blend, achado que contraria nossa hipótese inicial. Esse achado, por outro lado, ilumina uma interpretação do fenômeno no sentido de se combinarem boa frequência e preservação de segmentos fônicos para garantir que o blend seja interpretado adequadamente.

Conhecemos as limitações desta análise, empreendida sob a pretensão de se somar às descrições existentes sobre o tema. Há muitos desafios analíticos na análise de blends nas línguas do mundo e, de modo particular, em PB. Entre esses desafios, destacamos a necessidade de se precisar ou de se redimensionar a distinção entre blends fonológicos e morfológicos. Com base em Marangoni Jr. (2021), consideramos a semelhança fônica apenas no ponto de encontro das bases, ou seja, blends do tipo *escorleão* (*escorpião* + *leão*) foram classificados como morfológicos, ainda que suas bases possuam segmentos em comum, o que sinaliza para que se analise em pormenor, antes e para além de seu ponto de encontro, a semelhança fônica entre as palavras-base que

constituem o blend. Também blends com estrutura PP (ex. morcegonha < morcego + cegonha), cujo apagamento poderia em princípio ser atribuído a qualquer uma das duas bases, representam desafio. A ampliação de dados e de suas fontes e o aprofundamento do debate sobre frequência lexical e processamento psicolinguístico de blends se projetam como caminhos para melhor compreensão do fenômeno e de seus domínios, o objetivo maior de nossa investigação.

Agradecimento

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processos PQ-309576/2022-9 e PIBIC-310921/2018-0, e à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), processo PROBIC- 22/2551-0000763-5, pelo fomento sob forma de bolsas. Agradecemos à colega Isabela Prisco Petry pela contribuição em diversos estágios desta pesquisa.

Contribuição dos Autores

Este artigo corresponde a um subprojeto sobre blending inserido no projeto maior denominado *Domínios em Morfofonologia*, concebido e coordenado pelo primeiro autor. Além da construção conjunta da redação, segundo e terceiro autores participaram das etapas de levantamento e análise dos dados.

Referências

- ALVES, I. M. Neologismo. Criação lexical. 1. ed. São Paulo: Ática, 1990. 95p.
- ARAÚJO, G. A. Morfologia não-concatenativa em português: os portmanteaux. Caderno de Estudos Linguísticos, Campinas, v. 39, p. 5-21, 2000. DOI: <https://doi.org/10.20396/cel.v39i0.8636935>
- ARAUJO, L. C. de. Bichos Malucos. Jogos e materiais para alfabetização: oficinas, recursos e estratégias didáticas para a alfabetização, blogspot.com, 2012. Disponível em: <http://oficinasdealfabetizacao.blogspot.com/2012/01/bichos-malucos.html>. Acesso em: 18 mar. 2023.
- BASÍLIO, M. Cruzamentos vocabulares: o fator humorfológico. In: XII Congresso da ASSEL-RIO, 2003. Rio de Janeiro, 2003.
- BASÍLIO, M. Cruzamentos vocabulares como construções morfológicas. In: Congresso internacional da ABRALIN, 2005. Brasília. Anais do IV congresso internacional da ABRALIN, Brasília, 2005. 1600 p.
- BAT-EL, O. Blend. In: BROWN, K. (org.) Encyclopedia of Language & Linguistics. 2ª ed. Oxford: Elsevier, 2006, p. 66-70.
- BYBEE, J. Phonology and Language Use. Cambridge: Cambridge University Press, 2001.

- BYBEE, J. *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.
- CAMARA JR., J. M. *Estrutura da língua portuguesa*. Rio de Janeiro: Editora Vozes, 1970.
- CARVALHO, B.; BUENO, R. *Misturichos*. 1. ed. Brasil: WMF Martins Fontes, 2012. 32 p.
- GONÇALVES, C. A. Blends lexicais em português: não-concatenatividade e correspondência. *Veredas – Revista de Estudos Linguísticos*, Juiz de Fora, v. 7, n. 1, p. 149-167, 2003. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/25272/14296>. Acesso em: 20 dez. 2022.
- GRIES, S. Isn't that fantabulous?: How similarity motivates intentional morphological blends in English. In: ACHARD, M.; KEMMER, S. (eds.). *Language, Culture, and Mind*. Stanford: CSLI, 2004, p. 415-428.
- GRIES, S. Ambimorfemia de cruzamentos vocabulares: uma abordagem por ranking de restrições. *Revista da ABRALIN*, v. 5, n. 1/2, 2006. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/945>. Acesso em: 20 dez. 2022.
- GRIES, S.; ASSUNÇÃO, F. P. de. A humorfologia dos cruzamentos vocabulares em Português: análise da coluna de Agamenon, de "o Globo". *Veredas on-line - atemática*, Juiz de Fora, p. 57-71, 2009. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/25161/14192>. Acesso em: 17 abr. 2023.
- JUHASZ, B. J.; JOHNSON, R. L.; BREWER, J. An investigation into the processing of lexicalized English blend words: evidence from lexical decision and eye movements during reading. *Journal of Psycholinguistic Research*, v. 46, p. 281-294, 2017.
- KELLY, M. H. 'To "brunch" or to "brench"; Some Aspects of Blend Structure'. *Linguistics*, v. 36, n. 3, p. 579-590, 1998.
- LEHRER, A. Identifying and interpreting blends: An experimental approach. *Cognitive Linguistics*, v. 7, p. 359-390, 1996.
- MARANGONI JR., C. E. *A blendividade na formação de palavras: a derivação dos blends na interface entre morfologia, fonologia e pragmática*. 2021. Dissertação (Mestrado em Semiótica e Linguística Geral) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2021.
- MINUSSI, R. D.; NÓBREGA, V. A. A interface sintaxe-pragmática na formação de palavras: avaliando os pontos de acesso da Enciclopédia na arquitetura da gramática. *Veredas - Revista de Estudos Linguísticos*, Juiz de Fora, v. 18, n. 1, p. 161-184, 2014. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/24978/14010>. Acesso em: 03 dez. 2022.
- MINUSSI, R. D.; VILLALVA, A. Reconhecimento e acesso lexical dos blends em português europeu e português brasileiro. *Todas as Letras - Revista de Língua e Literatura*, v. 22, n. 1, p. 1-14, 2020. Disponível em: <https://editorarevistas.mackenzie.br/index.php/tl/article/view/12839/10538>. Acesso em: 03 dez. 2022.
- NÓBREGA, V. A.; MINUSSI, R. D. O tratamento da morfologia não-concatenativa pela morfologia não-concatenativa: o caso dos blends fonológicos. *Revista Letras*, Curitiba, n. 91, p. 158-177, 2015. Disponível em: <https://revistas.ufpr.br/letras/article/view/40094>. Acesso em: 03 dez. 2022.
- PIÑEROS, C. E. Word-blending as a case of non-concatenative morphology in Spanish. *Rutgers Optimality Archive*, 2000. Disponível em: <http://roa.rutgers.edu/files/343-0999/343-PINEROS-0-0.PDF>. Acesso em: 09 dez. 2022.
- PIÑEROS, C. E. The creation of portmanteaus in the extragrammatical morphology of Spanish. *Probus*, v. 16, p. 203-240, 2004.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2023. Disponível em: <https://www.R-project.org/>. Acesso em: 03 jun. 2023.

RSTUDIO TEAM. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA, 2023. Disponível em: <http://www.rstudio.com/>. Acesso em: 03 jun. 2023.

SANDALO, M. F. S. Morfologia. In: MUSSALIN, F.; BENTES, A. C. (Orgs.). Introdução à linguística: domínios e fronteiras. São Paulo: Cortez, 2005. p. 181-206.

SANDMANN, A. J. Formação de Palavras no Português Brasileiro Contemporâneo. Curitiba: Scientia et Labor, 1988.

SANDMANN, A. J. Morfologia Geral. São Paulo: Contexto, 1991.

SATRAPI, M. Ajidar, o dragão da terra. 2. ed. Brasil: Biruta, 2007. 32 p.

VILLALVA, A.; MINUSSI, R. D. Description and analysis of a Portuguese blend corpus. Corpus, v. 23, 2022. Disponível em: <http://journals.openedition.org/corpus/6436>. Acesso em: 02 ago. 2024.

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

WINTER, B. Statistics for linguists: an introduction using R. Taylor & Francis Group, 2020.