



MARCAÇÃO DE GÊNERO GRAMATICAL EM FORMAÇÕES NOVAS EM PORTUGUÊS BRASILEIRO

GRAMMATICAL GENDER MARKING ON NEW
FORMATIONS IN BRAZILIAN PORTUGUESE

Pedro Surreaux¹

Universidade Federal do Rio Grande do Sul

Luiz Carlos Schwindt²

Universidade Federal do Rio Grande do Sul

Resumo: Este trabalho investiga os fatores envolvidos na atribuição de gênero gramatical a novos substantivos em português brasileiro, assumindo o papel do conhecimento internalizado dos falantes quanto a regras e padrões presentes na língua. Para isso, analisam-se empréstimos introduzidos a partir do inglês, bem como dados de teste em que falantes nativos atribuem gênero a pseudopalavras. Os resultados mostram que, dentre os empréstimos do inglês, prevalece atribuição por analogia semântica, sendo terminações relevantes para o fenômeno restringidas pela fonotática da língua fonte; no caso das pseudopalavras com terminações associadas ao feminino, atribui-se feminino na maioria dos casos. Os resultados sugerem sistematicidade na atribuição de gênero gramatical a palavras novas, com a atuação de restrições de diferentes naturezas.

Palavras-Chave: Gênero Gramatical; Atribuição de Gênero; Empréstimos; Pseudopalavras; Morfofonologia.

¹ surreauxpp@gmail.com.

² schwindt@ufrgs.br. Bolsista de Produtividade em Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Abstract: *This work investigates factors involved in the assignment of grammatical gender to new nouns in Brazilian Portuguese, assuming the role of speakers' internalized knowledge concerning rules and patterns of the native language. We analyzed loanwords borrowed from English to Portuguese, as well as data from an experiment where speakers had to assign gender to pseudowords. The results show that, in the case of English loanwords, semantic analogy prevails as a criterion for gender assignment, due to the phonotactics of English, which inhibits phonological endings associated to gender in Portuguese. Among pseudowords, formal criteria appeared as relevant, where endings related to feminine have been assigned feminine in the tests. Results suggest systemacity in the assignment of gender to new words, where constraints of different natures interact.*

Keywords: Grammatical Gender; Gender Assignment; Loanwords; Pseudowords; Morphophonology.

INTRODUÇÃO

O gênero gramatical integra a morfologia nominal do português, expressando-se nos nomes e em palavras a eles relacionadas. Essa categoria manifesta-se de formas diferentes ao se contrastarem substantivos e demais itens nominais. Se desconsideramos o segmento final das palavras como um indicador intrínseco de gênero, a categoria só se torna observável através da sintaxe (ROCHA, 1998), no processo de concordância nominal (ex. a casa/o mapa, a tribo/o bolo, a ponte/o pente). Nessa operação, determinantes — artigos, adjetivos e alguns pronomes — são flexionados para gênero e número de acordo com valores contidos no substantivo, que atua como núcleo do sintagma nominal. Assim, considerando-se seu uso, um dado substantivo deverá apresentar, em relação ao gênero, uma de suas duas expressões possíveis na língua, masculino ou feminino. Entende-se, então, a natureza morfossintática do gênero gramatical.

Línguas que possuem gênero gramatical, de modo geral, correlacionam expressões desse sistema a traços que integram outros níveis da gramática. De acordo com a classificação de Corbett (1991), que considera a natureza e a expressividade dessas associações nessas línguas, o gênero do português poderia ser considerado um sistema híbrido, correlacionando-se em certo grau tanto à fonologia — em que se relaciona o segmento final -a e o gênero feminino —,

quanto à semântica, dada a relação do feminino semântico ao feminino gramatical.

Neste trabalho, investigamos o sistema de gênero do português tendo em vista essas interações, tomando como foco seu papel no sistema de atribuição de gênero quando da adaptação morfológica de itens novos na língua. Especificamente, buscamos melhor entender o que condiciona a escolha por um ou outro gênero quando substantivos novos são processados pelos falantes e sujeitos à gramática do português. Por exemplo, o que dá conta do gênero feminino atribuído à palavra *homepage* ou masculino à palavra *match*? Ou de uma expressão vernácula, como o substantivo *grama* (unidade de medida), que é frequentemente empregada no feminino apesar de registrada como masculina nos dicionários e gramáticas? Para isso, analisamos dados — de natureza aparentemente diversa — que permitem se observar a natureza regular do fenômeno em português brasileiro (PB), conforme sinaliza Corbett (1991, p. 70)³: os empréstimos nominais do inglês e os resultados de experimento de atribuição de gênero a pseudosubstantivos por critérios formais, realizado com falantes nativos para fins do presente estudo.

1 GÊNERO NA GRAMÁTICA E NO LÉXICO DO PB

Nesta seção, revisamos alguns estudos que colaboraram para a descrição e explicação de sistemas de gênero gramatical e suas interações com outros componentes da gramática. Inicialmente, abordamos estudos que se ocupam de formalização das categorias de gênero e classe temática na gramática do português, sendo eles Câmara Jr. (1970), Alcântara (2010) — com alguma

³ No capítulo que trata do “estatuto psicolinguístico da atribuição de gênero” (*the psycholinguistic status of gender assignment*, p. 70-104, tradução nossa), o autor cita o caso dos empréstimos nominais e dos experimentos com substantivos artificiais como tipos de evidência que ajudariam a explicar tal regularidade enquanto parte da competência do falante nativo.

incursão na análise do fenômeno em espanhol na perspectiva de Harris (1991), que fundamenta o trabalho de Alcântara —, Armelin (2014) e Schwindt (2011, 2018). Em seguida, oferecemos um breve panorama da produtividade da marcação de gênero em PB, baseando-nos especialmente em Schwindt (2018, 2020).

Ainda que nosso estudo não tenha por objetivo oferecer uma proposta de modelagem gramatical para o fenômeno, entendemos que sua melhor abordagem passa necessariamente pelo conhecimento das análises que descrevemos aqui. A razão é o entendimento de que a atribuição produtiva de gênero é processo que responde a demanda não de um mas de diferentes componentes linguísticos, tarefa com a qual os estudos que resenhamos procuram lidar.

1.1 Gênero e classe como categorias formais

Um dos primeiros autores a abordar o fenômeno do gênero gramatical em português foi Joaquim Mattoso Câmara Jr., na obra *Estrutura da Língua Portuguesa*, de 1970. Partindo de uma perspectiva estruturalista de Item-e-Arranjo, o autor entende o sistema de gênero da língua como comportando uma forma marcada, o feminino, que se opõe à forma não marcada, o masculino — tese que se baseia na observação dos segmentos finais dos itens nominais do português. As contrapartes femininas de pares de nomes sexuais, como *menino/menina* e *escritor/escritora*, seriam exemplares de uma instância de flexão nominal, operada pela anexação do morfema de gênero *-a* às formas femininas e pela ausência de marcas nas formas masculinas. Tal sistema de oposições se assemelharia ao de atribuição de número em português, em que o plural é instanciado pelo morfema *-s*, em oposição ao singular, representado por \emptyset . O restante dos nomes da língua, segundo essa perspectiva, uma vez que não

instanciam marcação de gênero, apresentam finais fonológicos relacionados, na morfologia, com a categoria de classe formal, que divide os nomes entre aqueles com tema em -a, -e, -o, ou atemáticos (vocábulos terminados em vogal acentuada).

A análise de Harris (1991) para o espanhol alimenta a análise proposta por Alcântara (2010) para o português. Harris analisa o sistema nominal do espanhol para explicar a exponenciação dos morfemas de gênero naquela língua, pressupondo a natureza modular das propriedades que subjazem substantivos, adjetivos, verbos e pronomes. Segundo o autor, substantivos carregam diferentes traços linguísticos, que podem, por sua vez, estar relacionados às categorias de gênero (feminino, masculino, biforme ou comum de dois gêneros), às classes formais (identificadas pelos finais fonológicos -o, -a, -e) e às propriedades semânticas (animado, humano, sexuado etc.). Ao lado dos nomes que pertencem ao que chama de *inner core*, aqueles que apresentam correlações mais regulares — itens masculinos que terminam em -o e femininos que terminam em -a —, estão casos com correlações gradativamente menos regulares — nomes femininos fechados por -o e masculinos fechados por -a.

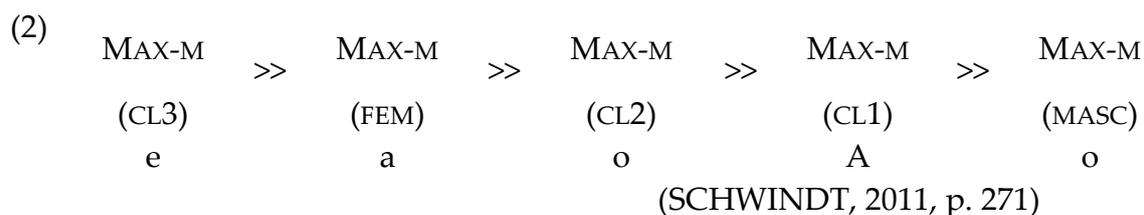
Alcântara (2010), em perspectiva semelhante, assume que nomes em português podem pertencer a uma de quatro classes formais: I, final em /o/; II, final em /a/; III, final em /e/; e IV, definida pela ausência de um elemento vocálico terminal, ou tema. Essas classes, assim como na proposta de Harris (1991), são potencialmente independentes em relação às duas classes de gênero dos nomes — feminino e masculino —, uma vez que podem ser observados na língua casos contemplando todas as possíveis intersecções entre classes formais e de gênero, ex. *ponte/pente, casa/mapa, tribo/livro*. A partir dessa classificação, Alcântara desenvolve uma análise com base na Morfologia Distribuída (HALLE; MARANTZ, 1993; 1994) para explicar a relação entre gênero e classe formal em português. A seleção de um dos três morfemas na derivação dos nomes é

atribuída a informações contidas nas raízes. Uma vez que o gênero masculino é o não marcado, ou *default*, cabe ao feminino uma informação específica, ou traço idiossincrático, a ser representado nas raízes, que seleciona a classe formal II na derivação, fazendo com que esta não precise estar explicitada, como acontecem em casos em que palavras masculinas terminam em *-a* (ex. *cometa*). Há, portanto, uma correlação inerente do gênero feminino com a terminação *-a*, prevista por regra, enquanto que a correlação dessa terminação com o masculino parte de especificação. Ainda, os casos raros em que femininos terminam em *-o* são explicados pela presença do traço de classe formal I, junto a *f*, na raiz, impedindo que o morfema II seja selecionado pela regra de redundância. Ou seja, é necessário que itens como *tribo* carreguem a informação específica quanto ao segmento terminal *-o*, de modo a impedir a aplicação da regra que relacionaria, de outro modo, itens femininos à vogal átona final *-a*.

Armelin (2014) distancia-se da perspectiva modular a respeito das classes nominais e oferece um tratamento unificado ao problema, entendendo gênero e classe temática como expoentes de um mesmo nó sintático na derivação. Ao contrário de Alcântara (2010), essa perspectiva não considera valores internos à raiz, que permitiriam a projeção ou seleção de complementos. Assim, é o nó sintático de gênero que seleciona a raiz. A derivação, nos casos menos marcados, por *default*, exponenciaria os segmentos terminais *-o* e *-a* no caso de raízes selecionadas por nós de gênero contendo, respectivamente, os traços masculino e feminino. Formas que fogem do padrão, como *mapa* e *cometa*, de um lado, e *tribo* e *libido*, de outro, seriam explicadas por especificações nos nós sintáticos que selecionam as raízes desses itens e atribuem segmentos terminais irregulares na inserção fonológica.

Em Schwindt (2011, 2018), temos uma perspectiva otimalista realizacional (MCCARTHY, 2007; 2011; WOLF, 2008) para dar conta dos nomes em português. Nessa análise, morfemas de gênero e classe temática são unidades abstratas

independentes que se exponenciam nos nomes com base em uma hierarquia que envolve restrições responsáveis pela realização de morfemas e restrições de marcação fonológica. Os candidatos em competição contemplam diferentes possibilidades de escansão de morfemas, e, inclusive, de sua não realização, caso de morfe zero em morfologia tradicional. A análise se concentra em nomes temáticos (fechados por *-o*, *-a* e *-e*), ainda que não se limite a esses nomes. À exceção do segmento final *-e*, que corresponde exclusivamente a um morfema de classe temática (CL3), cada final fonológico dos nomes temáticos (*-a* e *-o*) está potencialmente associado a dois morfemas, um de gênero e outro de classe formal (respectivamente: CL1 e FEM; CL2 e MASC) sendo, portanto, apesar de homófonas, unidades distintas do ponto de vista morfológico. A hierarquia é a que segue.



Em (2), CL3 aparece como o morfema mais marcado, o que está de acordo com os dados de produtividade descritos em Schwindt (2018, 2020): os substantivos em português dividem-se de modo equilibrado entre nomes femininos e masculinos no montante de palavras fechadas por *-e* átono; além disso, nomes sexuados fechados por *-e* não superficializam FEM a exemplo de *(o) estudante / (a) estudante*. A segunda posição, ocupada por FEM, justifica-se pela marcação de nomes femininos sexuados em *-a*, o que não acontece, à exceção de casos regulares *(aluno/aluna)*, com nomes masculinos ex., *professor/professora*. A maior produtividade de itens *impredizíveis* masculinos (terminados em *-a*, como *mapa* e *sistema*) em relação a itens femininos desse tipo (terminados em *-o*, como *tribo* e *libido*), sustentada pelos dados levantados em *corpora* envolvendo formas

dicionarizadas e dados de fala, embasa o ordenamento dos morfemas de classe temática CL2 e CL1. Por fim, a posição baixa na hierarquia do morfema MASC, substanciado por /o/, explica, por um lado, a agramaticalidade de palavras como "linguisto" para a maior parte dos falantes do PB, mas também a eventual emergência, na fala das crianças experimentando esse *ranking*, por exemplo, de formas como *formigo*.

1.2 Produtividade da marcação de gênero

Dadas essas abordagens sobre o sistema de gênero em português, que apontam para regularidades entre gênero e outros componentes da gramática, tratamos agora da produtividade da marcação de gênero. Neste trabalho, abordamos produtividade em duas perspectivas, a da quantificação das formas existentes no léxico e no uso de uma língua e a da potencialidade de criação de novas estruturas. Nessa direção, ainda em caráter de revisão da literatura, apresentamos a seguir dados quantificados sobre o emprego de gênero em PB, considerando as descrições de Schwindt (2018, 2020).

A descrição de Schwindt (2018, 2020) vale-se de duas amostras: 17.049 *types* do Dicionário Aurélio Eletrônico e 4.800 *tokens*/1.266 *types* de dados de língua falada do Projeto Variação Linguística na Região Sul do Brasil (VARSUL).

⁴ São substantivos, excluindo-se, para razões de foco e recorte, compostos e elementos de locuções.

De modo geral, nas duas amostras, há mais nomes femininos do que masculinos. Quanto à tipologia, há predomínio de uniformes, superando 90% dos dados, com ligeira preferência para femininos no uso, alavancada por palavras derivadas pelo sufixo -ção, inerentemente feminino na língua (ex. atração, empolgação). Biformes são menos de 3% no dicionário, número que se

⁴ <http://www.varsul.org.br>

eleva para 8% no uso. O número reduzido residual de nomes é de comuns de dois gêneros. No que respeita ao segmento terminal, em torno de 80% dos nomes das duas amostras são fechados pelas vogais átonas -a, -o, -e; dos restantes aproximadamente 10% são fechados por ditongos e outros 10% por consoantes e vogais tônicas. No que diz respeito aos nomes terminados em vogal átona, foco da discussão que fazemos aqui, enquanto a correspondência entre o segmento final -o e o masculino pode ser considerada categórica, a correspondência entre -a e feminino é de aproximadamente 95% nos dados do dicionário e de 90% nos dados do VARSUL.⁵ Palavras terminadas em -e, por outro lado, distribuem-se de forma equilibrada entre masculinas e femininas nas duas amostras. No subconjunto de nomes biformes que opõem -a/-o em uso, o equilíbrio entre nomes masculinos e femininos se desfaz, revelando-se a preferência por nomes masculinos. Inversa e curiosamente, no âmbito dos nomes fechados por -e, femininos superam importantemente masculinos. Entre os traços semânticos, é correspondência com sexo que se mostra o mais revelador na morfologia de gênero da língua. Apesar de haver poucos itens nesse grupo (5,5% no dicionário e 13,4% o VARSUL), sua saliência cognitiva associada ao significativo grau de isomorfismo entre palavras terminadas em -a e -o e, respectivamente, os gêneros feminino e masculino é possivelmente o que responde pela percepção de predizibilidade de gênero em PB segundo a intuição dos falantes. Outros traços, tais como caráter humano, comestibilidade ou mesmo concretude ou animacidade parecem exercer pouco ou nenhum papel na determinação de gênero no PB, ao contrário do que pode ocorrer em outras línguas.

⁵ A única palavra que constitui exceção importante para a relação -o/masculino em PB é *tribo*. O dicionário lista também *imago* e *virago*, palavras não usuais na língua, e *libido*, que parece apresentar gênero variável no uso. Também formas femininas deadjetivais podem assumir gênero masculino, como na construção *ela é uma soprano* (uma cantora do tipo soprano), o que escapa ao pressuposto de gênero como categoria básica dos substantivos que assumimos nesta análise.

Conforme Corbett (1991), sistemas de gênero gramatical, de modo geral, podem ser classificados como semânticos, formais, ou ambos, em menor ou maior grau. A produtividade do gênero em português aqui descrita aponta para um sistema híbrido: parcialmente formal, por correlacionar, por um lado, segmentos terminais a gênero, e parcialmente semântico, por parecer codificar na gramática traços semânticos de referentes sexuais.

2 A ADAPTAÇÃO DE SUBSTANTIVOS NOVOS A SISTEMAS DE GÊNERO

O contato linguístico é responsável por uma variedade de fenômenos nas línguas do mundo, sobretudo em nível fonológico, mas também envolvendo os níveis morfológico, sintático e semântico-pragmático. O tema deste artigo é a atribuição de gênero a itens novos em PB, que pode ser observada no fenômeno dos empréstimos lexicais, quando há a implementação de novas palavras em uma língua receptora a partir de uma língua fonte (SANKOFF, 2001). Ao ser integrado a um novo sistema, um dado empréstimo lexical naturalmente se adapta a suas estruturas e torna-se sujeito a seus processos gramaticais (HAUGEN, 1956).

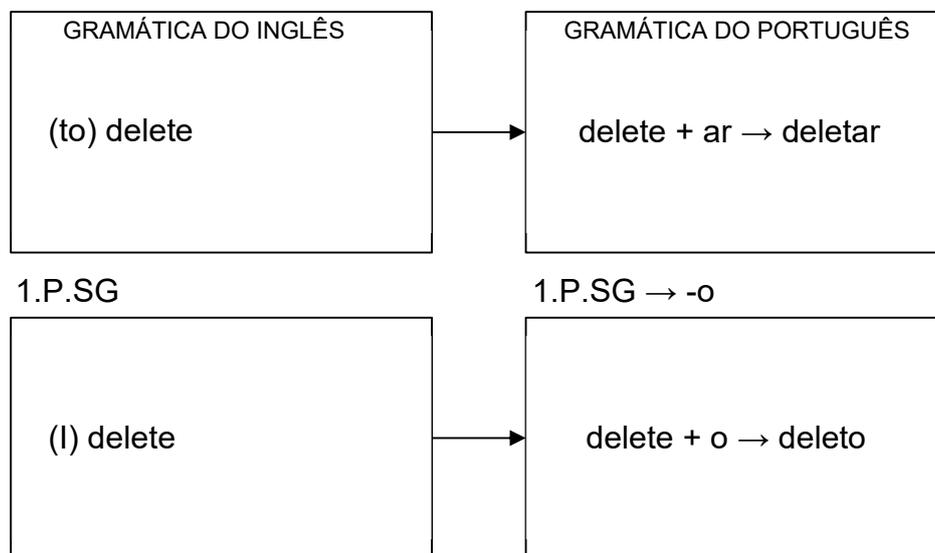
Entendemos empréstimos lexicais (do inglês, *loanwords*) de acordo com a definição de Haspelmath (2009). Empréstimos lexicais são palavras que passam a fazer parte de uma língua, em algum ponto de sua história, a partir de outra língua.⁶ Segundo o autor, as formas de origem dos empréstimos podem ser morfológicamente complexas ou mesmo frasais, mas perdem sua estrutura

⁶ Há uma questão terminológica em relação aos termos *borrowing* e *loanword*, recorrentes na literatura, ambos correspondendo a *empréstimo*, em português. Enquanto *borrowing* se refere ao processo de empréstimo, não só de itens lexicais, mas de estruturas e padrões prosódicos, fonológicos, sintáticos etc., *loanword* designa especificamente o produto do processo de empréstimo lexical (HASPELMATH, 2009, p. 38-39).

interna ao serem implementadas no novo sistema, onde são, em princípio, não analisáveis.

O processo de adaptação gramatical por que passam os empréstimos lexicais, e que os leva à estabilidade dentro do sistema receptor, implica, muitas vezes, a atribuição de características morfossintáticas ausentes na língua de origem. Esse fenômeno pode ser observado, por exemplo, em empréstimos verbais do inglês no PB. Para que possam funcionar na gramática da língua receptora, atribuem-se aos empréstimos os morfemas produtivos, abertamente expressos do paradigma verbal da língua, conforme a Figura 1.

Figura 1: Adaptação morfológica do verbo *deletar* em português



Fonte: Surreaux (2020, p. 24).

Ao ser adaptada à morfologia do português, a forma verbal *delete* pode ser expressa por *deletar*, com a afixação do sufixo infinitivo *-ar*, e, por decorrência, a todas as demais formas associadas no paradigma flexional desse verbo. Da mesma forma, empréstimos nominais em português apresentam as potencialidades morfológicas das classes nominais a que são alocadas, substantivos, em geral, apresentando flexão de número (*mouse / mouses*), de

gênero, para o caso dos comuns de dois gêneros (*a influencer / o influencer*), ou um traço de gênero inerente, em substantivos uniformes (*o link / a jukebox*).

O fenômeno da adaptação de empréstimos lexicais interessa a investigações de cunho fonológico, porque permite observar a emergência de padrões inerentes ao sistema receptor, sua análise interessa também à morfologia, por permitir que se observem não só produtos da aplicação de regras, mas também os critérios que condicionam a escolha por dadas formas em detrimento de outras, como é o caso da atribuição de gênero a empréstimos nominais.

Corbett (1991), ao abordar o fenômeno de atribuição de gênero gramatical a novos nomes, afirma que alguns sistemas baseiam-se predominantemente em critérios formais — como os segmentos terminais dos itens —, enquanto outros baseiam-se predominantemente em correlações com o componente semântico.

No caso dos critérios formais, em geral estão em jogo casos em que segmentos fonológicos de itens novos correlacionam-se com o gênero atribuído. Corbett (1991) exemplifica esse tipo de atribuição apresentando o caso do idioma haúça, da família afro-asiática, falada na Nigéria. Nessa língua, palavras do gênero feminino são categoricamente terminadas em *-aa*, regularidade que se estende a substantivos novos.

Critérios semânticos envolvidos na atribuição de gênero, segundo Corbett (1991), podem ser entendidos sob a generalização da associação conceitual. Para o autor, esse mecanismo manifesta-se em todas as línguas naturais, uma vez que integra a cognição humana, e seria responsável por agrupar conjuntos de palavras sob uma mesma classe de gênero morfológico com base em propriedades semânticas de seus referentes.

Como exposto acima, no português, bem como em grande parte das línguas indo-europeias, observa-se o papel das noções de sexo e gênero social do referente na categorização dos nomes que os designam. Nota-se, ainda, que

mesmo essa seleção de propriedades semânticas pela gramática é de certo modo arbitrária, uma vez que é específica a cada sistema. Esse fato torna-se evidente quando consideramos a diversidade de línguas que contam com categorias nominais análogas. Corbett (1991) ilustra essa diversidade com exemplos de línguas cujos sistemas de gênero são sensíveis a diferentes propriedades conceituais. Uma dessas propriedades é a edibilidade, atestada, por exemplo, no dyirbal, idioma falado na Austrália, em que uma das classes nominais inclui o conjunto de substantivos que designam frutas e vegetais comestíveis; outra propriedade é a instrumentalidade, observada no archi, língua falada no nordeste do Cáucaso, em que uma das classes nominais inclui palavras que designam ferramentas, armas e roupas; nomes também podem designar propriedades divinas, como ocorre no tamil, língua falada majoritariamente na Índia e no Sri Lanka, em que nomes que designam deuses ou entidades divinas associam-se ao masculino. Para além da influência de traços do referente na atribuição de gênero a empréstimos nominais, Corbett (1991) discute o critério de analogia semântica, segundo o qual novos itens recebem o gênero de palavras associadas presentes no léxico da língua receptora, como a atribuição do feminino no russo ao empréstimo francês *avenue*, possivelmente por uma analogia com o nome feminino *ulica* 'rua'. O autor conclui que esse critério deve ser entendido nas mesmas bases da associação conceitual. Thornton (2009) aprofunda o debate sobre o critério de analogia semântica, entendendo-o como uma categoria maior de fatores que atuam na atribuição de gênero gramatical. Ao observar conjuntos lexicais do italiano, a autora propõe os critérios de analogia com i) *hiperônimo*, ii) *conceito equivalente* e iii) *conceito associado*. Na atribuição pelo critério i), o gênero do substantivo que designa uma categoria é projetado em um empréstimo hipônimo; assim, em português, *uma yamaha* designa uma motocicleta (feminino), enquanto que *um yamaha* refere-se a um instrumento musical (um teclado, masculino). De acordo com o critério ii), o item

novo denota um conceito já existente na língua receptora e recebe o gênero por sinonímia, como nos empréstimos femininos *bike* e *night* e seus sinônimos, também femininos, *bicicleta* e *noite*. A atribuição pelo critério iii) se dá frente à ausência do conceito emprestado no léxico receptor, quando o empréstimo recebe o gênero de um substantivo associado semanticamente; como em *touchscreen*, feminina, em português por associação com *tela*.

3 EMPRÉSTIMOS DO INGLÊS EM PORTUGUÊS BRASILEIRO

Considerando-se o que afirma Corbett (1991) sobre o fenômeno da atribuição de gênero gramatical a novos itens e a possibilidade de sua observação no fenômeno dos empréstimos, apresentamos nesta seção o caso dos substantivos implementados no PB a partir do inglês, que respondem por um conjunto significativo dos substantivos novos da língua, e cujo processo de empréstimo está ainda em atividade.

O inglês compartilha as raízes indo-europeias do português e de outras línguas que, hoje em dia, contam com sistemas de gênero nominal. Essa língua, no entanto, passou por processo de empobrecimento morfológico em relação à categoria em etapas anteriores do idioma⁷. No inglês moderno, o gênero não se apresenta como sistema morfossintático, embora ainda se atestem, residualmente, no sistema pronominal da língua, a expressão de gênero semântico pelas formas pronominais pessoais de terceira pessoa *he* e *she*, *ele* e *ela*.

O conjunto de substantivos que analisamos neste estudo constitui, segundo classificação de Haugen (1956), casos de *loanword*, em que se conservam significado e forma de origem, descartando-se formas já completamente

⁷ Já no inglês antigo, a língua não contava mais com as expressões de gênero nos substantivos, que antes distinguiam feminino, masculino e neutro, e que perduram, por exemplo, no alemão moderno.

familiares ao sistema. A ausência de expressão de gênero nos substantivos do inglês nos dá segurança para excluir a hipótese de atribuição por analogia com o gênero dos itens na língua doadora, a que estão sujeitos empréstimos do francês e do espanhol, por exemplo.

3.1 Metodologia

Nosso *corpus* foi composto, inicialmente, pelos *corpora* de Cano e Prado (2006), Orsi (2015) e Valadares (2014), que reúnem empréstimos do inglês em PB referentes ao domínio da informática, da moda e de textos jornalísticos, respectivamente. Dada à contínua implementação desses itens à língua, passamos a ampliar o *corpus* com substantivos levantados por observações empíricas, verificando-se seu uso efetivo na língua por suas ocorrências no Corpus do Português: Web/Dialects, que conta com mais de um bilhão de palavras retiradas de páginas da *web*, com atualizações mensais.

O *corpus* totalizou 242 substantivos, dos quais 183 são uniformes, sendo 131 masculinos (ex. *notebook, hamburger*) e 52 femininos (ex. *webcam, picape*), e 17 são nomes comuns de dois gêneros (ex. *hacker, influencer*).

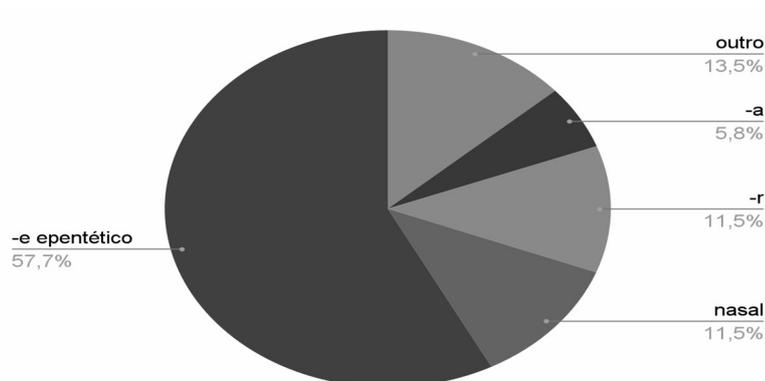
As variáveis analisadas foram *segmento terminal* (terminação em *-a, -e* epentético, *-er, -l, -s, -ks*, etc.) e *tipo* (substantivo feminino, masculino ou comum de dois gêneros). Tomando como pressuposto a ideia de Câmara Jr. (1970), corroborada por Schwindt (2011, 2018), de que o masculino é forma não marcada de gênero em português, centramos nossa análise nos 52 itens femininos, os quais foram codificados também de acordo com a variável *critério para atribuição de gênero feminino*, seguindo os critérios propostos por Thornton (2009) referentes a analogia semântica com itens femininos já presentes no léxico do português. Assim, poderiam atuar no processo a analogia com um *conceito associado* (ex. *ecobag*, em associação à *bolsa* ou *sacola*), com um *hiperônimo* (ex. *legging*, como

hipônimo de *calça* do tipo *legging*), ou um *conceito equivalente* (ex. *homepage*, como um equivalente tradutório de *página inicial*, em domínios da internet). Para além dos critérios de analogia semântica, estipulamos a presença da terminação átona final *-a* – ou grafema <*a*> – nas formas da língua fonte de empréstimos femininos como indicativa da atuação de um critério formal na atribuição de gênero.

3.2 Resultados

Conforme se vê na Figura 2, em relação à variável segmento terminal, os empréstimos femininos de nossa amostra se apresentaram, em ordem decrescente de frequência, como segue: *-e* final epentético (ex. *trademark*, *picape*); consoantes nasais (ex. *touchscreen*, *webcam*); *-r* (ex. *newsletter*, *happy hour*), *-a* (ex. *creepypasta*, *mídia*); outros segmentos (ex. *jukebox*, *skill*, *fake news*). Notamos ainda que 47% dos 17 itens comuns de dois gêneros apresentavam a terminação *-er*, do sufixo agentivo do inglês (ex. *hater*, *gamer*, *youtuber*). Esses itens, quando utilizados no feminino, seguem o padrão do funcionamento dos itens comuns de dois gêneros em português, fazendo referência ao sexo ou gênero do referente, como em *'a youtuber'*, designando referente feminino.

Figura 2: Segmento final de empréstimos femininos do inglês



Fonte: Surreaux (2020, p. 31).

Quanto aos critérios semânticos possivelmente atuantes na atribuição de gênero aos empréstimos, como mostra a Tabela 1, os 52 itens femininos distribuíram-se entre aqueles condicionados por analogia, sendo 54% de casos de conceito associado, 25% de equivalente tradutório, 7% de hiperônimos e 5,8% (3 casos) de presença do segmento final *a* na forma de origem (Tabela 1). Não foi possível relacionar 3 dos itens aos critérios de atribuição considerados: *gangue*, *van* e *performance*. A dificuldade de resgate desses itens deve-se possivelmente ao fato de terem sido implementados no português em um período anterior ao da maioria dos demais itens.

Tabela 1: Critério de atribuição de gênero a itens femininos do inglês

Critérios	Exemplos	Ocorrências
Critério formal (-a→fem.)	<i>mídia, creepypasta, sinuca</i>	3
Item associado no inventário	<i>selfie</i> (=foto), <i>playlist</i> (=lista)	28
Equivalente no inventário	<i>night</i> (=noite), <i>timeline</i> (=linha do tempo)	13
Hiperônimo no inventário	<i>legging</i> (=calça), <i>scooter</i> (=motocicleta)	5
Nenhum critério	<i>gangue, van, performance</i> ⁸	3

Fonte: Surreaux (2020, p. 31).

⁸ O reconhecimento sincrônico de palavras como *gangue*, *van* e *performance* como empréstimos é bastante discutível. Optamos, contudo, por incluí-las nesta análise considerando seu registro nas fontes consultadas. A não identificação de critérios para a atribuição de gênero a esses itens pode, porém, ser justamente sintoma de sua opacidade enquanto empréstimos.

3.3 Discussão

Devido à fonotática do inglês, predominam no recorte a atribuição de gênero por critérios semânticos, a partir da analogia com itens já presentes no inventário. Ou seja, a pouca ocorrência na língua fonte de itens terminados em *-a* ou outros segmentos que possam ser associados ao feminino em português faz com que a forma desses empréstimos seja pouco relevante para a escolha na maioria dos casos.

Dentre os três critérios semânticos, a prevalência por condicionamento por item associado no inventário (ex. *smartphone*, *fakenews*), bem como o condicionamento por hiperônimo (*legging*, *scooter*), atestado em menor escala, relacionam-se com a mencionada adoção de novos conceitos no contexto de globalização. A escolha por equivalente no inventário (*night*, *bike*), por outro lado, parece envolver fatores de natureza estilística e valor atribuído à língua fonte.

Nos três itens para os quais propomos condicionamento por critério formal, fica clara a natureza diversa dos processos que resultaram na presença do segmento vocálico final condicionante. A palavra *mídia*, por exemplo, entra no inglês através do latim, como *media*, plural de *medium* ('meio'); embora estranha ao padrão fonotático dessa língua, manteve-se a vogal átona final. *Creepypasta* deriva do portmanteau *coppypasta* (que, por sua vez, não figurou como empréstimo em português), usado em comunidades virtuais para referir contos que são difundidos ao serem copiados (*copy*) e colados (*paste*) sucessivamente⁹. A terminação *-pasta*, nesse caso, está relacionada ao registro na grafia de um processo de apagamento do segmento final *-d* no inglês e de abaixamento da vogal final resultante, a partir do substantivo de participial *coppypasted*. É semelhante o caso de *sinuca*, com origem na forma do inglês *snooker*, possivelmente associado ao feminino em português frente a sua pronúncia com

⁹ <https://time.com/2818192/creepypasta-coppypasta-slender-man/>. Acessado em 24/10/2020.

apagamento do *-r* final e abaixamento da vogal precedente: /'snu.ka/ (com a variante plenamente aportuguesada *sinuca*).

4 EXPERIMENTO COM PSEUDOPALAVRAS

Similarmente ao que ocorre com os empréstimos, um dado valor de gênero pode, segundo Corbett (1991), ser atribuído a substantivos novos com base em critérios formais quando a escolha do falante considera segmentos fonológicos, grafêmicos, ou padrões morfológicos da forma processada. Em português, isso pode ser exemplificado pela atribuição de feminino a uma palavra por conta de ser terminada por *-a*, marca de gênero associada estatisticamente ao feminino na língua. Com o objetivo de testar critérios fonológicos para atribuição de gênero a novos substantivos em PB, desenvolvemos um teste com pseudopalavras para testar a intuição de falantes nativos sobre o fenômeno.

4.1 Metodologia

Para a elaboração deste experimento, tomou-se como base os *corpora* do Dicionário Eletrônico Aurélio e do Projeto VARSUL, de onde foram levantados padrões de associação entre a forma dos substantivos e o gênero feminino. Há uma série de segmentos finais recorrentes na língua que estão estatisticamente relacionados ao gênero feminino, dentre eles as estruturas mórnicas que designamos aqui como sufixos *-(ç)ão*, *-agem*, *-ade*, *-ite*, *-ose*, e o segmento átono *e*. Aplicamos às pseudopalavras elaboradas tais segmentos a fim de testar o papel dessas correlações presentes no léxico sobre o fenômeno de atribuição de gênero a novos substantivos pelo falante.

O experimento contava com 60 questões, 40 das quais eram distratoras, elaboradas com base na fonotática do português, mas terminando em segmentos

infrequentes em palavras femininas, como *o, l, r e s*. As questões eram dispostas em ordem aleatória a cada vez que o teste era iniciado, a fim de neutralizar a influência do cansaço e da automatização nos resultados. Os participantes foram expostos a frases com as palavras-teste, cada uma das quais trazia uma pseudopalavra precedida por uma lacuna, que deveria ser preenchida com um dos dois artigos definidos da língua, tornando explícito o gênero atribuído a cada substantivo. O experimento contou, por exemplo, com a frase. *Já sabemos muito sobre ___ gutação*. Assumimos que os respondentes preencheriam a lacuna com <o>, se processassem a palavra como um substantivo masculino, ou <a>, se a entendessem como feminina. Em cada questão, era permitida a escolha de apenas um dos dois artigos, e o acesso às questões subsequentes só era possível uma vez respondidas as questões anteriores.

O requisito para a realização do experimento era que o participante fosse falante nativo de PB. O questionário iniciava com questões relativas às variáveis sociais *idade, sexo, escolaridade, cursos de graduação* (opcional/se aplicável) e *línguas faladas além do português* (opcional/se aplicável). O experimento foi elaborado e veiculado por meio da plataforma Google Forms e difundido através de redes sociais. Encerrado o experimento, os dados coletados foram convertidos em planilhas no Google Sheets e posteriormente analisados fazendo-se uso do programa SPSS Statistics (versão 25.0).

4.2 Resultados

O experimento contou com 89 respondentes. Dentre as terminações testadas, à exceção de *e*, todas foram empregadas predominantemente no feminino. Na Tabela 2, estão expostas as frequências com que o feminino foi atribuído a cada conjunto de pseudopalavras de mesma terminação, com todos

os itens testados, os quais não apresentaram diferenças estatísticas significativas no interior de cada grupo.

Tabela 2: Atribuição de feminino às terminações testadas

Terminação	Feminino	Pseudopalavras
-agem	98,50%	volinagem, ravidagem, delitagem
-dade	96,90%	sagueidade, lenvidade, bameridade
-ão	96,32%	gutação, estrização, arredição, ceisão
-a	93,42%	nantarra, jamência, passina, ralufata
-ite	76,80%	renutite, zemalite, debasite
-e	18,86%	filate, troje, dalome

Fonte: Autores.

Promoveu-se o cruzamento das variáveis sociais — *idade, sexo, escolaridade, cursos de graduação e línguas faladas além do português* — com as variáveis linguísticas em análise, que não se mostraram, em princípio, relevantes para a atribuição de gênero às palavras testadas.

4.3 Discussão

É possível se admitir que o processamento de novos itens, no que diz respeito à atribuição de valores de gênero, baseia-se, de alguma forma, em associações presentes no léxico da língua receptora. Os fatores formais testados no experimento apontam para um alto índice de atribuição de feminino a palavras novas com os finais *-agem, -dade, -ão, -a* e *-ite*, terminações bastante frequentes em palavras femininas do português. Nesse sentido, destaca-se o sufixo *-agem*, e em particular a pseudopalavra *delitagem*, que teve o feminino atribuído em 100% das ocorrências.

As três terminações com mais respostas femininas no teste apresentam correspondência nos *corpora* consultados. O sufixo *-dade* é categoricamente

feminino em ambos; *-agem* apresenta alguma variação em palavras comuns de dois gêneros (ex. a/o personagem), ao passo que *-ão*, ocorre em palavras masculinas uniformes e em palavras biformes (ex. chão; charlatão), o que pode explicar em alguma medida o menor índice de atribuição do feminino a esse segmento no experimento, como mostra a Tabela 2. Da mesma forma, o menor grau de atribuição de feminino pelos participantes às pseudopalavras terminadas em *-ite*, pode estar relacionada à homofonia desse sufixo com o segmento terminal de substantivos masculinos de uso frequente (ex. convite, limite).

As pseudopalavras com final *e*, fora de sufixo, tiveram o masculino atribuído em mais de 80% dos casos, contrastando com os dados dos *corpora* consultados, que apontam para uma distribuição equilibrada desse segmento entre os dois gêneros. Esse equilíbrio identificado nos *corpora* consultados, no entanto, não leva em consideração a presença desse segmento em sufixos. Se excluímos desses dados palavras terminadas em *-dade*, *-ite* e *-ete*, como apresentamos para as pseudopalavras na Tabela 2, chegamos a 28% de palavras femininas em *e* nos dados do dicionário, e, de 73%, passamos a 57% nos dados de fala. Assim, a correlação no léxico entre *e* e nomes masculinos, ao se excluirmos itens sufixados, pode servir para explicar a tendência observada quanto à atribuição do masculino a itens novos contendo essa terminação.

CONSIDERAÇÕES FINAIS

Conforme sugerido por trabalhos como Corbett (1991) e Poplack, Pousada e Sankoff (1982), a multiplicidade de fatores implicados no processo de atribuição de gênero a itens novos torna complexa a apreensão do processo. Ao analisarmos casos em que processos naturais geram contexto para o fenômeno, bem como casos em que tal contexto foi gerado artificialmente, constatamos a atuação de critérios de natureza formal e semântica, relacionados à produtividade de gênero

da língua e à distribuição das correlações relacionadas no léxico. A presença de índices como vogal final *a* e segmentos como *-agem* e *-dade*, predominantemente relacionados ao feminino no léxico, parece relevante para a atribuição de gênero; ainda, a analogia semântica com itens já presentes nos repertórios lexicais dos falantes pode projetar o gênero desses itens às novas palavras, conforme descrito pelos critérios de atribuição de Thornton (2009).

Propostas futuras de estudos sobre o tema incluem a abordagem de fatores de diferentes naturezas competindo em itens individuais, o que não pôde ser observado plenamente neste trabalho devido à ausência de fatores formais em inglês e à orientação específica do experimento elaborado. Além disso, também tendo em mente as limitações do inglês quanto à manifestação de fatores que não os semânticos, pode ser considerada a análise de recortes de empréstimos advindos de outras línguas em português.

REFERÊNCIAS

- ALCÂNTARA, C. As classes formais do português brasileiro. *Letras de Hoje*, Porto Alegre, v. 45, n. 1, p. 5–15, 2010.
- ARMELIN, P. Classifying nominals in Brazilian Portuguese: a unified account for gender and inflectional class. In: VESELOVSKÁ, L.; JANEBOVÁ, M. (orgs.). *Complex visibles out there: Proceedings of the Olomouc Linguistics Colloquium 2014: Language Use and Linguistic Structure*. Olomouc: Palacký University, 2014, p. 67-82.
- CÂMARA JR., J. M. *Estrutura da língua portuguesa*. 42. ed. Petrópolis, RJ: Vozes, 1970.
- CANO, W. M.; PRADO, D. de F. Os estrangeirismos da área da informática no Aurélio XXI. *Revista Alfa*, São Paulo, v. 50, n. 2, p. 265–275, 2006,
- CORBETT, G. *Gender*. Cambridge: Cambridge University Press, 1991.
- HALLE, M.; MARANTZ, A. Distributed Morphology and the Pieces of Inflection. In HALE, Kenneth; KEYSER (Eds.). *The View from Building 20*. Cambridge: MIT Press, 1993, p. 111-176.
- _____. Some key features of Distributed Morphology. *MIT Working Papers in Linguistics*, v. 21: Papers on phonology and morphology, p. 275-288, 1994.
- HARRIS, J. W. The exponence of gender in Spanish. *Linguistic Inquiry*, v. 22, n. 1, p. 27-62, 1991.

HASPELMATH, M. Lexical borrowing: Concepts and issues. In: HASPELMATH, M.; TADMOR, U. (Eds.). *Loanwords in the world's languages: a comparative handbook*. Berlin: Mouton de Gruyter, 2009. p. 35–54.

HAUGEN, E. *Bilingualism in the Americas: a bibliography and research guide*. Tuscaloosa: American Dialect Society, 1956.

MCCARTHY, J. J. *Hidden generalizations: phonological opacity in Optimality Theory*. London: Equinox, 2007.

_____. Pausal phonology and morpheme realization. In: BOROWSKY, Tony; KAWAHARA, Shigeto; SHINYA, Takahito; SUGAHARA, Mariko. *Prosody Matters: Essays in Honor of Lisa Selkirk*. London: Equinox, 2011.

ORSI, V. A presença de empréstimos da língua inglesa na revista brasileira Glamour. In: *Colóquio de Moda: 8ª Edição Internacional*, 11., 2015, Curitiba. *Anais eletrônicos [...]*. Curitiba:2015.Disponível em:<http://www.coloquiomoda.com.br/anais/Coloquio%20de%20Moda%20-%202015/COMUNICACAO-ORAL/CO-EIXO4-COMUNICACAO/CO-4-A-PRESENCA-DE-EMPRESTIMOS-DA-LINGUA-INGLESA-NA.pdf>. Acesso em: 17 jun. 2018.

POPLACK, S.; POUSADA, A; SANKOFF, D. Competing influences on gender assignment: variable process, stable outcome. *Lingua*, v. 57, p. 1–28, 1982.

ROCHA, L. C. *Estruturas morfológicas do português*. Belo Horizonte: Ed. da UFMG, 1998.

SANKOFF, G. Linguistic outcomes of language contact. In TRUDGILL, P., CHAMBERS, J.; SHILLING-ESTES, N. (Eds.). *Handbook of sociolinguistics*. Oxford: Basil Blackwell, 2001, p. 638-668.

SCHWINDT, L. C. Zeros na morfologia nominal portuguesa à luz da Optimal Interleaving Theory. *Revista virtual de estudos da linguagem – ReVEL*, n. 5, v. especial, p. 264-276, 2011.

_____. Exponência de gênero e classe temática em português brasileiro. *D.E.L.T.A.*, v. 34, n. 2, p. 745–768, 2018.

_____. Predizibilidade da marcação de gênero em substantivos no português brasileiro. In: DANNIEL CARVALHO, D.; BRITO, D. (orgs.). *Gênero e língua(gem): formas e usos*. Salvador: Editora da UFBA, 2020, p. 279-294.

SURREAUX, P. A atribuição de gênero gramatical a itens novos em português brasileiro. 2020. Trabalho de conclusão de graduação (Graduação) - Universidade Federal do Rio Grande do Sul. Instituto de Letras, Porto Alegre.

THORNTON, A. M. Constraining gender assignment rules. *Language Sciences*, v. 31, n. 1, p. 14–32, 2009.

VALADARES, F. B. Uso de estrangeirismos no Português Brasileiro: variação e mudança linguística. 2014. Tese (Doutorado em Letras). Pontifícia Universidade Católica de São Paulo/PUCSP, São Paulo.

WOLF, M. A. Optimal Interleaving: serial phonology-morphology interaction in a constraint-based model. 2008. Doctoral Dissertation. University of Massachusetts Amherst. ROA-996.

Nota do editor:

Artigo submetido para avaliação em: 27 de agosto de 2021.

Aprovado em sistema duplo cego em: 11 de outubro de 2021.