

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS

PEDRO PERINI SURREAUX

A ATRIBUIÇÃO DE GÊNERO GRAMATICAL A NOVAS FORMAÇÕES EM  
PORTUGUÊS BRASILEIRO

PORTO ALEGRE

2023

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS

PEDRO PERINI SURREAUX

A ATRIBUIÇÃO DE GÊNERO GRAMATICAL A NOVAS FORMAÇÕES EM  
PORTUGUÊS BRASILEIRO

Dissertação apresentada ao Programa de Pós-graduação do Instituto de Letras da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre em Teoria e Análise Linguística: Fonologia e Morfologia.

Orientador: Prof. Dr. Luiz Carlos Schwindt.

PORTO ALEGRE

2023

Dedico a meu querido tio e amigo Pedro Surreaux, que nos deixou enquanto se finalizava a escrita desta obra. Sua erudição, competência e sede por conhecimento me acompanharam nesta jornada, e seguirão elas e ele comigo.

## **AGRADECIMENTOS**

Agradeço à Universidade Federal do Rio Grande do Sul pela formação gratuita e de qualidade que me proporcionou.

À minha família, meu pai, Luiz, minha mãe, Márcia, meu tio, Pedro, meus irmãos e minha parceira Bárbara, que me apoiam e me permitem seguir trilhando este caminho.

Aos meus amigos de longa data, Afonso, Augusto, Leonardo, Alessandra, Eduardo, Venancio, entre outros que estiveram comigo em todos os momentos e que serviram de exemplo e inspiração.

A Júlia, Rossana, Camila, Isabela e Pedro Eugênio, meus amigos e colegas de pesquisa, com quem tanto aprendi, e que estiveram ao meu lado, dialogando e colaborando com suas experiências. Ao amigo Felipe Vital, que conheci através da linguística, e com quem tive trocas produtivas durante a escrita desta dissertação.

Aos professores e professoras do Instituto de Letras, Leandro, Valéria, Gabriel, Sérgio, Ingrid, que me inspiram e que me servem como exemplo de pesquisadores e profissionais da área da linguística.

Em especial, a meu orientador Luiz, que acreditou em mim e me abriu tantas portas e oportunidades, também servindo como enorme exemplo e inspiração para que eu trilhasse este caminho.

Ao grupo de estudos MorPhon, da UFRGS, por possibilitar tanto o diálogo com colegas como o contato com diversos textos interessantes da área dos estudos linguísticos.

E, não menos importante, a todos os participantes dos experimentos linguísticos referidos neste trabalho, de minha autoria ou não, que tiveram a disposição e o interesse em colaborar com o desenvolvimento da ciência.

“Por tudo isso, o gênero gramatical, à luz da linguística geral, é uma visão caleidoscópica, em que variam consideravelmente o número de classes, as linhas diretrizes da classificação e a maior ou menor coerências que essas linhas são obedecidas. [...] a categoria de gênero ilustra bem o que há de variável, complexo e convencionalmente arbitrário nas categorias gramaticais e como elas se distanciam das categorias lógicas.”

Mattoso Câmara

## RESUMO

Esta dissertação aborda o fenômeno de gênero gramatical e processos com que se relaciona, visando, primeiramente, a contribuir descritivamente para discussões sobre o tema. Toma-se como foco o sistema de gênero do português brasileiro e o processo de atribuição de gênero a substantivos novos na língua. Em português, todo substantivo em uso conta com um valor de gênero gramatical nuclear, uma vez que determina a flexão de seus determinantes, como artigos e adjetivos, na concordância nominal. Embora só seja possível se atestar o gênero de um substantivo pela sintaxe, no modo como se flexionam, por exemplo, artigos e adjetivos associados, há uma variedade de correlações entre o sistema de gênero e outros domínios linguísticos e extralinguísticos. Em PB, propriedades da forma fonológica dos nomes, como as terminações *a* e *agem*, podem indicar o feminino em muitos casos, assim como ocorre com substantivos que designam referentes do sexo feminino. Buscando apreender o fenômeno da atribuição de gênero a novos substantivos formados na língua, um objetivo inicial é reunir trabalhos relacionados ao tema, com uma revisão bibliográfica que se pretende abrangente, incluindo, de um lado, estudos que buscam caracterizar o gênero gramatical enquanto fato linguístico, de outro, investigações sobre o fenômeno da adaptação morfológica de substantivos a sistemas de gênero. A partir da compreensão que esperamos fornecer através desse compilado bibliográfico, partimos para a exposição de nossos empreendimentos metodológicos de análise *in loco*, em maior ou menor grau, do fenômeno; sejam eles: (i) a formação de empréstimos nominais em português derivados do inglês, do tupi e do árabe; (ii) e experimentos com falantes de português brasileiro utilizando-se experimento psicolinguístico com pseudopalavras, que deve testar critérios formais possivelmente atuantes no processo. Entendemos ter papel nesses fenômenos o conhecimento internalizado dos falantes quanto a regras e padrões lexicais do português. Os dados de nossas análises de *corpora* de empréstimo sugerem que os 1.150 substantivos formados a partir de empréstimos do inglês, em grande parte, adquirem gênero por processos de analogia semântica, como o sexo do referente ou uma palavra equivalente em português, uma vez que estruturas relevantes para o funcionamento do fenômeno são restringidas pela fonotática do inglês. Por outro lado, dentre os 827 empréstimos do árabe e os 2.544 de línguas tupi em português, apontamos para o fato de que substantivos femininos que terminados em *a*, em português, como

alcatra e capivara, já possuíam segmentos análogos em suas línguas de origem — nesses exemplos, <al.qatra> e <kapi'iûara>. Por fim, o experimento psicolinguístico online que realizamos com 133 nativos de PB apontou para um papel de radicais de palavras naturais do português na atribuição do feminino a palavras como brincadeirofe e televiséi, entendidas como femininas por 68,4 e 70% dos falantes, respectivamente. No entanto, o teste de terminações como ção e agem parece estar envolvidos de maneira mais incisiva no processamento de gênero gramatical: as pseudopalavras testadas nafiquência e matidade foram entendidas como femininas por 98,5 e 97% dos participantes, respectivamente. Nossa expectativa com as discussões e análises de dados empíricos aqui propostas é contribuir para pesquisas futuras sobre o gênero gramatical e elucidar o seu processamento enquanto categoria cognitiva. Por fim, ainda que de maneira mais tangencial, com nossa compilação e classificação específica de vocábulos de origens árabe e de línguas tupi, esperamos também oferecer possibilidades de estudo para a linguística histórica.

**Palavras-chave:** gênero gramatical, atribuição de gênero, português brasileiro, morfologia, empréstimos lexicais, pseudopalavras, psicolinguística.

## **ABSTRACT**

This thesis addresses the phenomenon of grammatical gender and processes related to it, aiming to contribute descriptively to discussions on the topic. The focus here is the gender system of Brazilian Portuguese and the process of gender assignment to novel nouns. In Portuguese, every noun has a nuclear grammatical gender value, thus determining the inflection of agreement-targets, such as articles and adjectives in nominal agreement. Although the gender value of a noun can only be attested in syntax, by the form assumed by determiners, there are certain correlations between gender and other linguistics and extralinguistic properties, such as semantic information of referents and given phonological constituents of the noun. The feminine gender in Portuguese is in some cases associated with semantic features and sex of referents and unstressed vowel ending *a*. In order to study the phenomenon of gender assignment to novel items in the language, the objectives are to carry out a comprehensive bibliographic review on the grammatical gender system of Brazilian Portuguese and on the phenomenon of morphological adaptation to gender systems; to study the phenomenon from nominal borrowings from English, Tupian, and Arabic into the language; and to conduct experiments with Brazilian Portuguese speakers using pseudonouns to test formal criteria that may be involved in the process. It is understood that internalized linguistic knowledge of speakers regarding rules and lexical patterns of within Portuguese lexicon may play a role in the phenomenon. According to our results, semantic analogy prevails in the case of English borrowing, once relevant formal clues for the phenomenon being restricted by English phonotactics; among borrowings from Arabic and Tupi, a significant part of the feminine nouns ending in *a* derives from words ending in analogous unstressed vowels of those languages. Regarding our psycholinguistic tests, pseudowords with predominantly feminine endings and roots in the Portuguese lexicon were very frequently assigned feminine. The results of our analyses point to the involvement of different clues as conditioning factors for gender assignment, resounding findings from the literature on other languages. Moreover, in the analyzed cases, there seems to be relative systematicity in the application of the process that goes beyond cases associated with the semantics of referents, sometimes assumed in the literature to be the only systematic and regular criterion.



**Keywords:** grammatical gender, Brazilian Portuguese, morphology, lexical borrowings, pseudowords, psycholinguistics.

## LISTA DE QUADROS E FIGURAS

|   |    |
|---|----|
| Quadro 1 – Modelo de complexidade de atribuição de gênero.....            | 34 |
| Figura 1 – Adição de Nó de Sufixo Temático a $X^0$ .....                  | 31 |
| Figura 2 – Derivação de nomes da Classe I.....                            | 31 |
| Figura 3 – Estrutura com SN como projeção máxima.....                     | 35 |
| Figura 4 – Estrutura de um SN sob a proposta da Decomposição Lexical..... | 36 |
| Figura 5 – Estrutura nominal sob um núcleo SD.....                        | 37 |

## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 1 – Atribuição de feminino com base nas terminações.....             | 70 |
| Tabela 2 – Atribuição de feminino com nas raízes.....                       | 71 |
| Tabela 3 – Atribuição de feminino e número de sílabas de R.nat+T.artif..... | 72 |
| Tabela 4 – Atribuição de feminino e número de sílabas de R.nat+T.artif..... | 72 |
| Tabela 5 – Gênero dos empréstimos do árabe.....                             | 77 |
| Tabela 6 – Terminação dos empréstimos do árabe.....                         | 77 |
| Tabela 7 – Gênero dos empréstimos do tupi.....                              | 82 |
| Tabela 8 – Terminações dos empréstimos do tupi.....                         | 82 |
| Tabela 9 – Cruzamento entre gênero e terminação.....                        | 83 |
| Tabela 10 – Gênero dos empréstimos do inglês.....                           | 86 |
| Tabela 11 – Segmento terminal dos empréstimos do inglês.....                | 87 |

## SUMÁRIO

|  |     |
|--|-----|
| 1 INTRODUÇÃO.....  | 13  |
| 2 O GÊNERO GRAMATICAL.....   | 17  |
| 2.1 Delimitando o fenômeno.....  | 17  |
| 2.1.1 O gênero nas gramáticas tradicionais.....                                  | 19  |
| 2.1.2 O gênero na linguística moderna.....                                       | 27  |
| 2.1.2.1 A abordagem estruturalista de Câmara Jr.....                             | 27  |
| 2.1.2.2 Abordagens gerativistas.....   | 28  |
| 2.1.2.2.1 O gênero como uma categoria sintática.....                             | 35  |
| 2.2 O gênero gramatical em processos psicolinguísticos.....                      | 42  |
| 2.2.1 Aquisição de sistemas de gênero.....                                       | 42  |
| 2.2.2 Experimentos psicolinguísticos.....  | 48  |
| 2.2.3 O gênero em empréstimos nominais.....                                      | 50  |
| 3 EXPERIMENTO COM PSEUDOPALAVRAS.....  | 64  |
| 3.1 Metodologia.....   | 65  |
| 3.2 Resultados.....  | 70  |
| 3.3 Discussão.....   | 72  |
| 4 ANÁLISE DE <i>CORPORA</i> DE EMPRÉSTIMOS.....                                  | 74  |
| 4.1 Empréstimos do árabe.....  | 75  |
| 4.2 Empréstimos do tupi.....   | 79  |
| 4.3 Empréstimos do inglês.....   | 85  |
| 4.4 Síntese do capítulo.....   | 88  |
| 5 CONSIDERAÇÕES FINAIS.....  | 90  |
| REFERÊNCIAS.....   | 92  |
| APÊNDICE.....  | 97  |
| A) <i>Corpus</i> de empréstimos do árabe.....                                    | 97  |
| B) <i>Corpus</i> de empréstimos do tupi.....                                     | 102 |
| C) <i>Corpus</i> de empréstimos do inglês.....                                   | 119 |
| D) Termo de Consentimento Livre Esclarecido do experimento psicolinguístico..... | 127 |
| E) Materiais do questionário.....  | 129 |
| F) Convite para a participação no experimento psicolinguístico.....              | 129 |

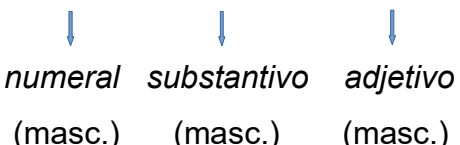
## 1 INTRODUÇÃO

Este trabalho tem como objetivo investigar a propriedade linguística de gênero e fenômenos a ele relacionados. Especificamente, busca-se compreender de forma mais aprofundada o processo em que falantes do português brasileiro<sup>1</sup> processam novos substantivos e lhes atribuem informação de gênero.

A noção de gênero gramatical como um fenômeno linguístico tende a variar, a depender da abordagem metodológica que se adota e do aspecto específico que se propõe a estudar. Em português, o gênero dos substantivos rege a forma expressa pelos demais vocábulos nominais:

(1)

Hoje encontrei dois escritores famosos.



*numeral*   *substantivo*   *adjetivo*  
 (masc.)   (masc.)   (masc.)

Se, de outro modo, o enunciado acima narrasse um encontro com *escritoras*, a forma flexionada dos dois determinantes seria feminina: *duas* e *famosas*. Entendemos então que a expressão de gênero de *dois* e *famosos*, em (1), é determinada pelo substantivo, que é central nesse processo de concordância. Entretanto, nos interessa saber como e de onde é atribuído o gênero dos substantivos, sendo essa a questão que norteia nossa investigação.

Tendo como foco o gênero em substantivos, apresentamos a seguir, de forma bastante sumarizada, algumas possíveis caracterizações do fenômeno comumente assumidas na literatura:

(2)

- (a) Entende-se como sistema classificatório que divide os substantivos em duas ou mais categorias de acordo com propriedades morfológicas e sintáticas — p. ex., concordância com artigos, adjetivos e verbos — ou semânticas, como a animacidade e o sexo dos referentes<sup>2</sup>;

<sup>1</sup> Por questões de praticidade e economia, doravante PB.

<sup>2</sup> Câmara Jr. (1964). Esta abordagem se assemelha àquela assumida pelas descrições de gramáticas tradicionais de modo geral (apresentadas a seguir, em 2.1.1.), mas também fundamenta propostas contemporâneas de modelos tipológicos (p.ex. FEDDEN & CORBETT, 2016).

- (b) Corresponde a um traço interno que é atribuído aos nomes em sua derivação sintática e garante sua boa formação na gramática e seu funcionamento em processos como a concordância<sup>3</sup>;
- (c) É uma forma pelas quais as línguas humanas refletem, na organização de suas estruturas, a categorização de elementos do mundo. Assim como o sexo dos seres animados para as línguas românicas, outras propriedades conceituais subjazer classificações nominais das línguas, como a animacidade, o caráter humano, a quantidade e diversas possibilidades de combinações entre classes<sup>4</sup>.

A existência de diferentes definições e aproximações possíveis para o fenômeno demonstra o caráter complexo e multifacetado do objeto de estudo desta dissertação. Um de nossos objetivos, em princípio, é que esta dissertação contribua para estudos futuros sobre gênero gramatical através da compilação bibliográfica apresentada no Capítulo 2 a seguir. Para isso, e também a fim de estabelecer nossas bases teóricas, buscamos oferecer, na Seção 2.1, uma revisão abrangente dos estudos sobre gênero, desde as primeiras descrições do sistema de gênero do português apresentadas por gramáticas históricas da língua, passando por sua abordagem na tradição gramatical contemporânea e chegando às análises estruturalistas e cognitivistas da linguística moderna.

Para além do gênero como propriedade gramatical e suas caracterizações, o processo de sua atribuição a novas formações nominais<sup>5</sup> é central para as discussões aqui propostas. Assim como outros processos cognitivos, o processamento de informações linguísticas ocorre, em geral, abaixo do nível da consciência<sup>6</sup>. Ao ouvirmos ou lermos uma palavra nova, por exemplo um substantivo, assumimos, sem maiores esforços conscientes, uma série de propriedades que lhes são inerentes e que garantem seu funcionamento na língua.

<sup>3</sup> Pressuposto de teorias gerativistas como a Morfologia Distribuída (HALLE & MARANTZ, 1993), que entendem gênero como um dos traços da matriz morfossintática dos nomes (abordado em detalhe em 2.1.2.2.1).

<sup>4</sup> Concepção que não dá conta da arbitrariedade atestada em muitos substantivos, em especial os inanimados, quanto a categorias conceituais, segundo Câmara Jr. (1964, p. 130–131).

<sup>5</sup> No campo dos estudos dos empréstimos lexicais, definem-se como substantivos que são emprestados ou adotados de outra língua para a língua receptora (*novel nouns*, conforme Corbett, 1991). Ainda que se estudem mais comumente do ponto de vista sincrônico, a caracterização ampla do fenômeno abre espaço também para análises da introdução de substantivos também na diacronia, como aquelas aqui apresentadas no Capítulo 4.

<sup>6</sup> Hyönä, J., & Lehtonen, P. (2010) argumentam que uma grande parte do processamento da linguagem ocorre de forma inconsciente, e que nossos cérebros utilizam processos automáticos para atribuir características gramaticais às palavras, como os substantivos.

Sabemos de forma implícita que: (i) podemos fazer uso de um sufixo -s para informar a existência de dois ou mais de seus referentes; (ii) seu emprego nos enunciados se associa mais diretamente a palavras específicas, como os pronomes, os adjetivos e os determinantes, que com ele concordam mas também lhe conferem certas informações semânticas; e, principalmente, (iii) deve ser empregado em um gênero específico.

Tendo em vista que o foco do nosso estudo é o fato (iii) do processamento de novas formações, assumimos uma concepção de gênero gramatical que se aproxima àquela em (1.b). Desse modo, entendemos que uma abordagem como a da Morfologia Distribuída<sup>7</sup> é a que melhor daria conta desse fato linguístico em específico; mais especificamente, consideramos em nossa argumentação as propostas apresentadas em Kramer (2015, 2016), Scher e Malagoli (2022) e Burkholder (2018) quanto ao traço de gênero e sua localidade na estrutura dos nominais.

Tendo isso em vista trazemos na Seção 2.2 uma revisão de trabalhos apresentando, em diferentes línguas, casos em que se pode atestar a atribuição de gênero a novos substantivos: a aquisição linguística (Item 2.2.1), a realização de experimentos psicolinguísticos (Item 2.2.2) e a formação de empréstimos lexicais (2.2.3). Com isso, objetivamos traçar um panorama das possibilidades de apreensão e descrição do fenômeno de atribuição de gênero gramatical a novas formações lexicais e, por fim, direcionar a discussão para a seção seguinte.

Na seção 3, apresentamos um experimento, realizado para esta investigação, que explora o papel de pistas formais, tanto de bases como de sufixos relacionados ao feminino, na atribuição de gênero pelos falantes. São expostos os procedimentos metodológicos, resultados e análise, e, por fim, são discutidos parcialmente os dados obtidos.

Nosso foco no Capítulo 4 é a análise de corpora (em relação a SURREAUX, 2020) acerca de empréstimos linguísticos do português vindos de diferentes línguas, reunidos em corpora, sobre os quais faremos a discussão adequada quanto ao fenômeno aqui estudado.

---

<sup>7</sup> Doravante, MD.

Em geral, este trabalho tem como objetivo contribuir para uma melhor compreensão de como o gênero é atribuído em português e como pode variar em diferentes contextos.



## 2 O GÊNERO GRAMATICAL

Abaixo, na Seção 2.1, trazemos uma revisão da literatura que enfoca o gênero gramatical enquanto fenômeno linguístico, desde suas abordagens em gramáticas mais antigas do português, como Soares Barbosa (1822) e Said Ali (1923), da gramática prescritiva contemporânea de Cunha e Cintra (2007), até análises de viés estruturalista e cognitivista. Na seção 2.2, discutimos em específico o fenômeno que é o tema desta dissertação, trazendo trabalhos dedicados à atribuição de gênero na aquisição linguística, em experimentos com pseudopalavras e em processos de empréstimo.

### 2.1 Delimitando o fenômeno

Em geral, línguas que possuem um sistema de gênero gramatical correlacionam suas propriedades a outras características da gramática. A exemplo do português, o gênero feminino, por exemplo, pode estar associado a sufixos (*-dade, -agem, -ção* etc.), ao sexo de um referente, (*a escritora, a influencer* etc.). De acordo com Corbett (1991), os sistemas de gênero das línguas do mundo podem ser altamente diversos, variando muito em termos do número de valores, tipos de critérios de atribuição e envolvimento em processos gramaticais. Algumas línguas têm apenas dois gêneros, enquanto outras têm mais de 20. Os critérios de atribuição de gênero podem incluir fatores como propriedades semânticas, fonológicas e morfológicas dos substantivos, bem como fatores sociais ou culturais. O gênero também pode variar em termos de padrões de concordância, onde e como é marcado, e se é marcado obrigatória ou opcionalmente em vários elementos dentro da gramática. Portanto, os sistemas de gênero podem apresentar uma ampla variedade de diversidade entre as línguas.

De acordo com a classificação de Corbett, o gênero do português pode ser considerado um sistema híbrido, pois se relaciona tanto à fonologia quanto à semântica. Na seção seguinte veremos algumas análises ao longo da história do português que indicam esta natureza da atribuição de gênero da língua.

Sabe-se que o sistema de gênero do português se manifesta na estrutura — seja externa ou interna — dos nominais. Esse traço se organiza, de modo geral, em um esquema binário que compreende masculino e feminino. A dupla potencialidade de expressão desses valores de gênero é característica dos nomes em francês,

espanhol, italiano, catalão, e na grande maioria das línguas românicas faladas atualmente.

O latim, ancestral comum mais próximo compartilhado por elas, no entanto, possuía um sistema de gênero tripartite, com feminino, masculino e um terceiro valor, chamado de gênero neutro. Na transição do latim para as línguas românicas, os substantivos neutros foram recategorizados como masculinos ou femininos, em grande parte por conta de suas diferentes formas, que seriam associadas a terminações ora de um, ora de outro gênero do sistema emergente.

Segundo Alkire e Rosen (2010) processo de perda do gênero neutro nas línguas românicas começou no latim popular. Este foi um período de grande mudança linguística, pois o latim era falado por pessoas de diferentes regiões com diferentes dialetos. Como resultado, havia muita variação na forma como os substantivos neutros eram usados. Alguns falantes começaram a reanalisar os substantivos neutros como masculinos ou femininos, e essa tendência se tornou generalizada com o tempo.

Ainda segundo esses autores, a perda das distinções de caso nas línguas românicas contribuiu para a perda do gênero neutro. Em latim, os substantivos neutros possuíam diferentes formas de caso, o que os ajudava a se distinguir dos substantivos masculinos e femininos. No entanto, à medida que o sistema de casos declinava nas línguas românicas, o gênero neutro se tornou menos distinto. Isso tornou mais fácil para os falantes simplesmente reanalisar os substantivos neutros como masculinos ou femininos. Por exemplo, a *vela*, forma plural do neutro *velum* deu origem a uma palavra feminina do português, possivelmente por analogia da vogal átona final *a* com o gênero feminino.

Hoje, apenas alguns traços do gênero neutro permanece nas línguas românicas. No caso dos substantivos, são tipicamente substantivos concretos que se referem a objetos inanimados, e aparecem no romeno. No caso do português, um resquício do neutro pode ser observado no paradigma dos pronomes demonstrativos, como *aquele* e *este*, ao lado de *isto*, *aquilo*, e *esta* e, *aquela*.

Essa diáspora do neutro nas línguas românicas é um exemplo de mudança linguística e também da ação de forças que reorganizam os sistemas morfológicos, nesse caso, um sistema de gênero é empobrecido, perdendo uma de suas classes, cujos membros são redistribuído redistribuídos nas restantes categorias. Este processo ainda está em andamento, de certo modo, na maneira como itens ainda

são recategorizados por aprendizes de línguas românicas como primeira e segunda língua, e também no modo como novos substantivos são percebidos pelos falantes em relação a seu gênero.

### 2.1.1 O gênero nas gramáticas tradicionais

Nesta seção, trazemos algumas das primeiras descrições gramaticais do sistema de gênero português que se propuseram a elencar propriedades dos nomes que estariam associadas ao masculino ou o feminino. Essas classificações nos são importantes porque apontam índices para regras de atribuição de gênero.

Trindade e Viaro (2022), apresentam a relação entre gênero gramatical e semântica em substantivos sexuais, que relaciona entidades masculinas e femininas a tipos de gênero gramatical de mesma nomenclatura, e a *arbitrariedade*, que estaria subjacente ao traço de gênero dos demais substantivos, uniformes quanto a gênero.

Essa arbitrariedade estaria evidenciada i) interlinguisticamente, na discrepância de gênero entre certos pares de substantivos cognatos em, por exemplo, português e espanhol, como *cor* (*f.*) e *color* (cor, em espanhol; *m.*); e ii) intralinguisticamente, no “gênero vacilante” de algumas dessas palavras, como *champagne*, que pode ser masculino ou feminino em português.

Os fenômenos referenciados por Trindade e Viaro para demonstrar a arbitrariedade do gênero de substantivos não sexuais em português são especialmente interessantes para nossa discussão. De um lado, o que é mencionado em i) diz respeito ao percurso diacrônico do léxico de sistemas que preservam a categoria gramatical de gênero, aqui discutido através dos substantivos latinos que se preservaram até variedades mais recentes do português e do espanhol. Por outro lado, o fato em ii) está relacionado ao período de adaptação em que, ao serem introduzidos no português e em sua gramática de gênero, certos substantivos não sexuais apresentam alguma variação quanto ao traço.

Antes de seguirmos para a apreciação de alguns autores da gramática tradicional quanto à categoria de gênero gramatical, apresentamos a seguir alguns dos problemas que aponta Câmara Jr. (1970) em relação a essas caracterizações. O primeiro ponto é o da incoerência ao se assumir o aspecto semântico do gênero das palavras. Não só é comum que muitos nomes em português façam referência a objetos desprovidos dos conceitos de sexo biológico e gênero identitário, mas

também certos substantivos sexuados são indiferentes a informações desse tipo, como *testemunha* e *cônjuge*.

A segunda crítica é direcionada a descrições do fenômeno de gênero que não diferenciam a flexão de processos lexicais ou sintáticos usado para indicar o sexo do referente. Para Câmara Jr. (1970), a flexão de gênero do português propriamente dita dedica-se a um número limitado de casos, instanciados, por exemplo, por *-a*, *-ã* e *-ona*, como em *gata*, *irmã* e *valentona*, que se opõem a *gato*, *irmão* e *valentão* (p. 90). No entanto, as gramáticas tradicionais generalizam, de forma incoerente, segundo ele, ao incluir nessa descrição os “nomes que variam em gênero por heteronímia”, como *homem* e *mulher*, os epicenos, que recorrem à sintaxe para fazer a oposição, como *cobra* (*cobra macho* e *cobra fêmea*), e os casos de derivação, que expressam o feminino por sufixos como *-triz* e *-esa*, em *imperatriz* e *baronesa*, por exemplo.

Em sua *Grammatica philosophica da lingua portugueza* (1774), tida como uma das mais importantes gramáticas do português<sup>8</sup> Jerónimo Soares Barbosa traz uma classificação dos substantivos quanto ao gênero baseada em dois fatores que seriam determinantes para essa expressão: *significação* e *terminação*. Quanto à *significação*, o gênero gramatical dos substantivos estaria associado ao sexo do referente, mas, nota-se, para alguns dos nomes femininos, também a certos domínios semânticos, como o das *virtudes*, das *paixões* e das *artes liberais*. Já para o fator *terminação*, o gramático português lista 43 terminações, 28 das quais estariam mais estreitamente relacionadas a um dos gêneros, e 15 que entende como *incertas* quanto a esse tipo de associação.

Em uma discussão posterior, Leoni (1838) parte do modelo classificatório de Barboza acerca dos fatores que explicariam a natureza do gênero dos substantivos, propondo agora aprofundamentos quanto à descrição das duas categorias. Para a relação dos gêneros com a *significação*, o autor acrescenta subcategorias também para os substantivos masculinos, como nomes de ventos e rios. Em relação ao fator *terminação*, Leoni contribui com uma discussão mais detalhada sobre as associações predominantes entre o gênero e a forma dos nomes, chamando atenção para os casos que escapam a essas regularidades. Leoni busca na etimologia explicações tanto para as associações mais regulares entre gêneros e

---

<sup>8</sup> Publicada no período do Iluminismo em Portugal, essa gramática contribuiu para a consolidação do estudo da língua portuguesa enquanto disciplina acadêmica.

terminações quanto para suas exceções, apontando a origem de alguns substantivos como a causa para a imprevisibilidade do gênero adotado.

Said Ali, em sua *Gramática Secundária* (1923), caracteriza o gênero gramatical do português primeiramente por sua manifestação morfossintática. Tomando os substantivos como o *locus* do fenômeno, o gramático os define como masculinos quando podem seguir o artigo *o*, ser combinados com qualificativos terminados em *-o* e substituídos pelo pronome *ele*; e como femininos quando podem seguir o artigo *a*, ser atribuídos aos qualificativos fechados por *-a* e substituídos por *ela*. Além disso, a gramática de Said Ali inclui a existência de pares de substantivos sexuais, epicenos e comuns de dois gêneros e o modo como o gênero gramatical neles se expressa na caracterização do fenômeno. A descrição de Said Ali (1923), desse modo, vai ao encontro das classificações em Barboza (1822) e Leoni (1838) ao adotar a divisão entre *significação* e *terminação*.

Na *Grammatica historica da lingua portugueza* (1931), Said Ali aprimora sua descrição, apresentando uma sistematização mais detalhada que a de seus predecessores, com mais índices formais e semânticos associados às expressões de gênero gramatical no léxico do português (numeração 1–4, abaixo). Alicerçando-se nas descrições das gramáticas anteriores mencionadas, Said Ali parte dos fatores *significação* e *terminação*, apresentando, para cada um, uma série de categorias de substantivos que expressam, de modo notavelmente regular, o feminino ou o masculino.

Para o gênero com base semântica, os grupos de nomes masculinos e femininos que designam seres sexuais, à exceção dos epicenos, são apresentados como categorias independentes na classificação de Said Ali. Ao lado delas, um número de outros domínios semânticos entendidos como proximamente associados a um ou outro gênero são elencados. Nota-se, entretanto, a distinção assumida pela atribuição de gênero baseada no sexo do referente e aquela que ocorre com base em conjuntos semânticos que independem desse traço. Ao contrário dos sexuais, os substantivos em (1b–e) e (2b), abaixo, apresentam, na perspectiva de Said Ali, valor de gênero meramente morfológico, arbitrário portanto.

## 1. Masculinos

a. seres sexuais masculinos (não epicenos): homem, rei, mestre, sacerdote, conde, boi, bode etc;

- b. pontos cardeais: norte, sul, oeste, oriente;
- c. letras, notas musicais e algarismos: o bê, o cê, o jota, o beta, o dó, o ré, o mi, o três, o cinco, etc;
- d. meses: março, abril, etc;
- e. rios, montes, mares e ventos: o Amazonas, o Himalaya, o Atlântico, etc.

## 2. Femininos:

- a. seres sexuados femininos (não epicenos): mulher, mãe, vaca, cabra, etc;
- b. nomes geográficos em que são subentendidas palavras femininas como ilha e cidade: Nova Friburgo, Ceylão, etc. (SAID ALI, 1923, p. 59–60)

Para a categorização que considera o fator *gênero pela terminação*, Said Ali atenta para sua maior adequação para a descrição dos “nomes de coisas” e dos substantivos sexuados epicenos, os quais, nota-se, eram a exceção às categorias em 1–a e 2–a. O gramático entende que, quanto ao fator *gênero pela terminação*, substantivos como *vítima* ou *criança*, embora denotem seres sexuados, estão sujeitos às mesmas regras do grupo dos não sexuados. Abaixo apresentamos a lista das terminações associadas ao masculino (3) e ao feminino (4), conforme elencadas pelo autor:

## 3. Masculinos

- a. -o átono
- b. -ema
- c. -oma
- d. -grama
- e. oxítonos em -á, -é, -i, -ó, -u
- f. terminados em ditongo puro
- g. oxítonos em -em, -im, -om
- h. -men átono
- i. termos científicos em -en átono
- j. -um
- k. concretos em -ão
- l. -r
- m. -l
- n. oxítonos em -az, -iz, -oz e -uz
- o. paroxítonos em -s e -x

- p. -ate, -ete, -ote, -ite
- q. concretos em -ude
- r. -arte e -orte
- s. -ante, -ente, onte

#### 4. Femininos:

- a. -a
- b. -dem e -gem átonos
- c. -ão, se abstratos
- d. -ade, -ude, -ice
- e. -ie
- f. -ede, -ide, -ave, -ebe, -eve
- g. -ase, -asse, -ace, -ese, -ece, -esse, -ose;
- h. além de exceções às regras do masculino

Nota-se que, embora tenha sido referido pelas três gramáticas como parte de *terminação*, Said Ali explicita o papel do acento nas caracterizações, um fator formal de outra natureza que interage com as terminações. Assim, as terminações em 4b só correspondem ao feminino se em posição átona, enquanto que as terminações em 3e devem ser acentuadas para que haja correlação com o masculino.

Por fim, destaca-se a interação entre critérios formais e semânticos presente na sistematização de Said Ali, um dos pontos que ilustra seu rigor descritivo e um maior grau de detalhamento em relação às classificações de gramáticos anteriores. Essa correlação entre índices formais e semânticos se observa nas categorias 3i, k e q, e 4–c, em que descrições de grupos formais trazem especificações de natureza semântica. Desse modo, delimitam-se substantivos masculinos com a terminação *en* átona ao domínio dos termos científicos e os terminados em *ão* e *ude* à designação de referentes concretos; e os femininos em *ão* ao domínio dos abstratos.

Aqui entendemos que a subordinação de fatores semânticos a formais, e não o contrário, está relacionada à maior abrangência destes em relação àqueles: as terminações agrupam um grande volume de palavras, enquanto os grupos semânticos propostos, à exceção de 1–a e 2–a, dizem respeito a conjuntos muito

menores. Nota-se ainda, quanto às especificações semânticas para palavras com uma mesma terminação, a possível atuação de processos derivacionais distintos mas com afixos homófonos. Em 4–c, que indica que substantivos abstratos em -ão são femininos, *precisão* e *lentidão*, por exemplo, formam-se a partir da regra em 5–a, enquanto que os masculinos em *ão*, como *valentão* e *telão*, são, em geral, aumentativos (5–b).

5)

a) Adj + -ão > N(abstr.)

b) N + -ão > N(concr.)

O tratamento oferecido pela tradição gramatical brasileira, ao se propor a descrever e prescrever as estruturas dessa variedade do português é realizada com algumas incoerências e de superficialidades, segundo aponta Câmara Jr. (1970). Um dos mais notórios compêndios gramaticais da língua traz definições, em grande parte, infundadas e imprecisas sobre aspectos que vão desde a breve definição introdutória de gênero, sob o título da seção a ele dedicada, até a exposição de seu comportamento e do lugar que ocupa no sistema de que é parte e com que interage.

Segundo Cunha e Cintra (2007), o gênero é uma das três flexões dos substantivos do português, ao lado de número e grau. Na seção dedicada ao fenômeno – *Gênero*, sob *Flexões dos substantivos*, no capítulo *SUBSTANTIVO* –, os autores assumem a existência de um gênero masculino, o qual chamam “termo não marcado”, e um feminino, “termo marcado” (p. 202), sem maiores explicações para o que entendem por essas noções. Como veremos na seção seguinte, essas noções foram possivelmente transpostas da análise de Câmara Jr. (1970) —, não triviais, em especial ao se considerar o público-alvo da obra.

Substantivos, segundo a gramática, não permitem a apreensão de seu gênero, seja por sua significação, seja por sua terminação, sendo que assumem como masculinos ou femininos todos os substantivos a que podem ser antepostos, respectivamente, os artigos *o* e *a*. Ainda que admitida tal imprevisibilidade, os autores apresentam extensa série de generalizações quanto à forma e o significado dos substantivos em relação a sua classificação de gênero, as quais entendem como sendo *convenientes à facilidade de aprendizado* (p. 203).

Quanto à significação, seriam “em geral” (p. 203) masculinos, nomes de 1a) homens e funções que desempenham; 1b) animais do sexo masculino; 1c) lagos,



montes, oceanos, rios e ventos; e 1d) meses e pontos cardeais. De modo similar, em vista do significado dos substantivos femininos, apresenta-se uma série de generalizações, segundo as quais seriam “normalmente femininos”, nomes de 2-a) mulheres e suas funções; 2-b) animais do sexo feminino; e 2-c) cidades e ilhas. Nota-se que, na forma de observações sob os títulos referentes aos itens acima enumerados, os autores propõem que os hiperônimos em 1c-d) e 2c) têm seu gênero subentendido na estrutura dos termos que categorizam (ex., o [*monte*] Everest, o [*oceano*] Pacífico, o [*mês de*] Março; e a antiga [*cidade*] Ouro Preto, as [*ilhas*] Antilhas, a [*ilha da*] Córseica). Essa proposta dos autores é interessante para a discussão sobre os critérios semânticos para atribuição de gênero, apresentada na seção 4.3 abaixo.

Quanto à terminação dos substantivos, os autores assumem categoricamente que são masculinos nomes com final átono -o, enquanto que aqueles terminados em -a são “geralmente” (p. 204) femininos. A terminação no ditongo nasal -ão estaria distribuída, via de regra, entre masculinos, para o conjunto de nomes concretos, e femininos, para substantivos abstratos, com a exceção, observam, da palavra *mão*, que, embora concreta, é feminina. O item dedicado às generalizações quanto a terminações encerra com a consideração de que, para além dos casos listados, “é sempre difícil se conhecer pela terminação o gênero de um dado substantivo.” (p. 204)

A seguir, é apresentada uma descrição do modo por que se associam pares de substantivos com referentes animados, que “costumam flexionar-se em gênero”, com uma forma para cada gênero. Nota-se, antes de apresentar sua descrição, o modo como são referidas as associações e unidades envolvidas na análise. É recorrente o uso de “desinência” para designar diversos segmentos terminais dos nomes, que se envolvem na “formação” de femininos a partir de masculinos, pelo “acréscimo” e “substituição” dessas terminações.

Assim, em uma categorização inicial, opõem-se pares em que a forma feminina é “completamente diversa” (p. 205) da masculina, ou seja, supletivos, de raízes distintas, e pares em que se assume uma raiz compartilhada entre as duas formas, onde a feminina “deriva” da masculina.

As chamadas “regras gerais” (p. 206) derivam formas femininas de masculinas pela 3.a) substituição de -o por -a e 3.b) acréscimo de -a sobre finais

consonantais. Em observação, os autores sugerem ainda que desinências especiais formam femininos pela 3.c) substituição do final -o por outras desinências.

Já a extensa descrição das “regras especiais” (pp. 206-209) traz o conjunto de pares de substantivos que seriam formados por 4.a) substituição de -ão por -oa, -ã e -ona; 4.b) acréscimo de -a a finais -or; 4.c) substituição de terminações por -esa, -essa e -isa; 4.d) substituição de e por a. Por fim o texto traz um grupo de pares, sob justificativa de serem “dignos de nota”, os quais apresentam variadas associações entre terminações: 4.e) substituição de ô por ó, a por isa, o por ina e acréscimo de esa, ina e isa a terminações consonantais ou em vogal tônica, entre outras.

A seção dedicada a gênero segue com a categorização dos substantivos que designam referentes animados entendidos como apresentando gênero uniforme, sendo eles os nomes 5.a) epicenos; 5.b) sobrecomuns; e 5.c) comuns de dois gêneros. Conclui-se a seção com observações sobre a origem dos nomes em 5.c), os quais incluiriam a totalidade dos substantivos deadjetivais terminados em -ista, e sobre o comportamento excepcional do substantivo *personagem*, o qual se poderia empregar em um ou outro gênero independente do referente.

Em suas partes finais, a seção elenca os casos em que 6.a) pares de substantivos de mesma forma variam a significação para um e outro gênero; 6.b) substantivos masculinos terminam em a — os quais se admitem numerosos, embora seja sustentado que “a terminação a seja de regra denotadora do feminino” (p. 211) —; e 6.c) substantivos de *gênero vacilante*, os quais são empregados em um ou outro gênero indiscriminadamente.

Para além dos empasses terminológicos quanto à sua conceituação, é seguro dizer que o gênero gramatical, em português, se relaciona com substantivos, adjetivos, artigos e pronomes pessoais. Enquanto que adjetivos e artigos dependem do substantivo e do adjetivo para manifestarem um ou outro gênero, substantivos, e por vezes, pronomes, possuem um desses valores, em geral, de forma intrínseca. É seguro afirmar que o gênero gramatical é um fenômeno essencialmente morfossintático, uma vez que a imprevisibilidade quanto ao valor de um dos itens acima citados só se desfaz no processo de concordância.

Assim, tomando como ponto de partida os empreendimentos descritivos da gramática tradicional do português, passamos agora para a revisão da literatura científica sobre o tema, que oferece diversas abordagens explicativas para os fatos

das distribuições de marcas e valores de gênero e os processos em que a categoria está implicada.

### 2.1.2 O gênero na linguística moderna

Nesta seção apresentamos discussões mais recentes quanto às propriedades do sistema de gênero em português, as quais abordam seus padrões de marcação e produtividade à luz do estruturalismo e de teorias de base gerativista. Iniciando pela análise Item-e-Arranjo de Mattoso Câmara (1970), revisamos os trabalhos de Harris (1991), Alcântara (2010) e Armelin (2014), os três sob a ótica da Morfologia Distribuída (HALLE; MARANTZ, 1993, 1994), e de Schwindt (2011, 2018), na perspectiva da Teoria da Otimidade (PRINCE; SMOLENSKY, 1991).

#### 2.1.2.1 A abordagem estruturalista de Câmara Jr.

Tendo inaugurado os estudos linguísticos em perspectiva estruturalista no Brasil, o estruturalista Mattoso Câmara Jr. aborda diversos aspectos da língua portuguesa em sua obra. Entre suas contribuições, está a breve mas robusta descrição do fenômeno do gênero gramatical português e seus padrões de marcação, em *Estrutura da Língua Portuguesa*, de 1970. Câmara Jr. propõe um sistema de oposição entre duas expressões possíveis de gênero no português, o masculino e o feminino. O estatuto de forma marcada do feminino tem como base padrões identificados nas terminações dos substantivos animados. Assim, em pares como professor/professora, é explícita a ausência de uma marca na forma masculina, em detrimento da marca *a* em sua contraparte, que indicaria, deste modo, o feminino. Já para pares em que a ausência de marcas não é explícita na parte masculina, como os pares masculino/feminino que opõem uma terminação em *a* a uma em *o* (gato/gata, onde o *o* em gato ainda pode ser visto como marca de masculino), Mattoso Câmara atribui às vogais finais da parte masculina o caráter de vogal temática, as quais corrigem formas que outrora terminaram em segmentos não atestados em português (como seria o caso da form *\*/gat/*).

Além disso, para Câmara Jr., as contrapartes femininas de pares de nomes sexuais, como menino/menina e escritor/escritora seriam exemplares de uma instância de flexão nominal, operada pela afixação de *-a* formadora de nomes animados femininos que designam entidades femininas, e que se assemelha à

flexão de número, em que o plural é instanciado pelo morfema -s. Segundo o autor, o caráter não marcado do masculino está evidente em outro recorte dos nomes sexuais: i) nos casos mais regulares, pares apresentam final *a* para feminino e *o* para masculino; ii) é possível que o nome feminino termine em *a* sem que a contraparte masculina traga elementos além da raiz, p. ex., professor/professora; iii) há casos em que há a ausência de marcas tanto no masculino como no feminino ex. (o) estudante/(a) estudante. A noção de marcação, neste caso, prevê que não existe a possibilidade de apenas a forma masculina de um par receber a marca de gênero: \*a professor/o professoro, \*a estudante/o estudanto. Além disso, seria evidente o esquema de marcação de gênero no modo com que a forma pluralizada de raízes que designam seres animados, como “alunos”, designa especificamente entidades femininas quando a marca de plural acompanha o feminino (“alunas”), mas inclusivo dos dois gêneros quando acompanha a flexão masculina.

As linhas esboçadas por Câmara Jr. (1970) sobre o gênero gramatical do português são influentes, como vimos para muito do que se falava e se fala sobre o fenômeno, estando inclusive presente nas gramáticas tradicionais, que trazem a noção de não marcação do masculino em relação ao feminino. Tal entendimento do sistema é crucial para as análises e discussões que serão apresentadas aqui, seja por assumirmos o masculino como tipo não marcado do sistema de gênero português para o processo de atribuição de gênero a novas formações, seja porque está representada nos alicerces de trabalhos que o seguiram, como os que resenhamos na sequência.

### 2.1.2.2 Abordagens gerativistas

A gramática de gênero do português assemelha-se em muito com aquela das demais línguas românicas ibéricas, tanto por seu sistema binário, que opõe masculino e feminino, como pelas marcas formais que caracterizam cada um dos gêneros, como -a para feminino. Devido à maior hegemonia do espanhol em relação ao português nos estudos linguísticos, apresentamos aqui uma investigação inicial de viés gerativista sobre o espanhol, que aborda a sistemática do gênero gramatical, mas que agrega para o estudo dessa categoria em português.

Harris (1991) analisa o sistema nominal do espanhol para explicar a exponenciação dos morfemas de gênero naquela língua. A análise pressupõe a natureza modular das propriedades que subjazem os substantivos, adjetivos, verbos

e pronomes. Segundo o autor, os 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.).

Uma das características da abordagem de Harris é o aspecto modular atribuído aos sistemas nominais: um nome pode ser feminino ou masculino, terminar em uma variedade de segmentos fonológicos, e relacionar-se a referentes que apresentam sexo biológico ou gênero social masculino ou feminino, embora nenhuma dessas classes categoriais esteja necessariamente correlacionada a gênero.

Harris (1991) classifica os nomes da língua de acordo com a regularidade das correlações entre os diferentes traços a eles atribuídos; assim, aqueles pertencentes ao que rotula como *inner core* seriam os exemplos mais regulares, onde itens masculinos terminam em *o* e femininos em *a*, seguido por um número de casos com correlações gradativamente menos regulares, sendo os últimos casos os mais raros: aqueles em que nomes femininos terminam em *o* e masculinos em *a*. Essa classificação, que considera a recorrência das correlações, é análoga em quase todos os aspectos ao que acontece no português, como veremos em trabalhos que, até certo ponto, baseiam-se em Harris para analisar o sistema de gênero desta língua.

Alcântara (2010) adota uma fundamentação que de um modo geral é análoga àquela de Harris ao classificar os morfemas nominais do português quanto a gênero e classe temática. Utilizaremos essa análise descritiva para elucidar o funcionamento do processo de derivação à luz da MD no que diz respeito às informações de gênero e classe temática.

A autora analisa a estrutura dos não verbos<sup>9</sup> do PB propondo uma categorização para os nominais não derivados que terminam uma das três vogais átonas *o*, *a* e *e* — p. ex. *calmo*, *tema* e *clube*, respectivamente — ou fechados por zero fonológico — como *troféu*, *jovem*, *chá* etc. As quatro categorias entendidas pela autora como *classes formais* são: classe I, das palavras terminadas em /o/; II,

---

<sup>9</sup> A autora retoma o recorte de seu estudo referido como aquele dos *vocábulos nominais*, limitando-se em sua análise, no entanto, aos substantivos e adjetivos.

das terminadas em /a/; III, associada aos nomes terminados em -e em suas formas do singular e àqueles onde se atesta essa vogal no plural, antes do morfema -s – como *juíz, juízes; lar, lares*; e IV, dos nomes atemáticos, cuja classe formal se associa ao morfema  $\emptyset$  e ao zero fonológico. Essas classes, assim com na análise de Harris, são em princípio independentes do gênero atribuído aos nomes, uma vez que podem ser observados na língua casos em todas as possíveis intersecções das classes formais e classes de gênero, mesmo que em diferentes níveis, ex. *ponte/pente, casa/mapa, tribol/livro*.

A partir desse modelo classificatório, Alcântara desenvolve sua análise para explicar a relação entre a gênero e classe formal em português. Esse tratamento teórico coloca o português ao lado de outras línguas românicas, como o espanhol e o italiano, no que diz respeito à exigência de um sufixo temático em não verbos para cumprir uma condição de boa-formação da língua. 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 na raiz. Assim, de modo similar à regra de redundância apresentada por Harris (1991), a informação de *f*, feminino, nas raízes, seleciona a classe formal II na derivação, fazendo com que esta não precise estar explicitada, como é o caso de masculinos terminados em *a* (p. ex., *cometa*).

Segundo a MD, unidades linguísticas são produzidas a partir de determinadas etapas, onde recebem as informações e estruturas que as constituem. Na dimensão inicial da derivação, a Sintaxe, determinadas unidades abstratas – como as raízes e os traços sintáticos, morfossintáticos e semânticos – são selecionadas e combinadas em estruturas a serem processadas a seguir pela morfologia e pela fonologia. Segundo Alcântara, as operações relevantes para sua análise dos nominais do PB são Inserção Tardia, Estrutura Sintática Total e Subespecificação<sup>10</sup>.

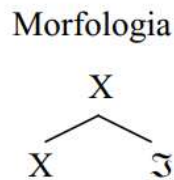
A Estrutura Sintática Total implicaria a adjunção de um morfema categorizador a raízes abstratas para a formação das estruturas nominais funcionais em PB. Conforme a representação em (7), a abordagem de Alcântara

---

<sup>10</sup> *Late Insertion, Syntactic Hierarchical Structure All the Way Down e Underspecification.*

postula para a vocábulos nominais não derivados ( $X^0$ ) como *pele* uma operação morfológica que adiciona um morfema de classe ( $\mathfrak{S}$ ) a raízes ( $\sqrt{\quad}$ ).

Figura 1 – Adição de Nó de Sufixo Temático a  $X^0$ .

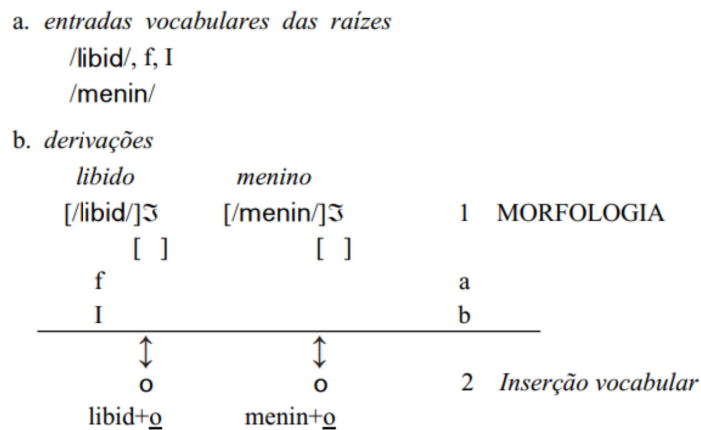


(Adaptado de ALCÂNTARA, 2010, p. 7)

A partir dessa estrutura morfológica bem formada, são inseridas entradas vocabulares em associação a cada um de seus nós terminais.

A Inserção Tardia relaciona-se ao caráter abstrato de estruturas sintáticas e morfológicas como raízes e morfemas de número e classe formal, que só adquirem substância fonológica ao final da derivação. Na Morfologia, a Inserção Vocabular opera sobre unidades abstratas como  $\mathfrak{S}$  e  $\sqrt{\quad}$ , atribuindo a elas entradas vocabulares, por sua vez relacionadas a seus correlatos fonológicos. Assim, o substantivo *barco*, por exemplo, contém: uma raiz abstrata e um morfema de classe formal, aos quais serão inseridas, na Fonologia, sequências /barc/ e /o/, respectivamente. Essas etapas da derivação estão representadas abaixo na Figura 8.

Figura 2 – Derivação de nomes da Classe I.



(ALCÂNTARA, 2010, p. 9)

Por sua vez, o conceito de Subespecificação, em MD, explica a atribuição de certas informações a unidades na derivação por meio de regras que não são explicitadas na gramática, o que se relaciona ao exemplo fortuito daquilo que assume Câmara Jr. (1964, 1970) quanto à não marcação do masculino, em sua abordagem sobre o gênero em português<sup>11</sup>. Sob a ótica da DM, relações de marcação e não marcação se estendem a outras propriedades gramaticais, como classe formal, onde costumam ser representadas por meio de Regras de Redundância<sup>12</sup>.

Partindo de pressupostos sobre o caráter *default* de certos traços e associações envolvidas na derivação de nomes em PB, Alcântara assume três casos em que se opera subespecificação (9): a) a presença do morfema de classe formal I, o qual não é informado na estrutura interna dos nomes, mas se atribui nos casos em que não há especificação para as demais classes I, II e III; b) a presença do traço de gênero masculino, instanciado pela ausência da marcação pelo feminino; e c) a informação quanto à classe formal II quando ocorre o traço de gênero feminino.

9)

1. [atribuição de CL I como default]
2. [atribuição de masc. como default]
3. [atribuição default de II a fem.]

A autora fundamenta as regras em 9) nas seguintes generalizações sobre o léxico nominal do PB. a) A maior parte dos vocábulos nominais termina em /o/. b) Nomes masculinos quase sempre terminam em /o/. c) Nomes femininos quase sempre terminam em /a/. Nota-se que, no esquema de derivação morfológica de Alcântara, gênero precede classe formal na ordem em que se processam os traços morfológicos das raízes, sendo ele quem determina as terminações, e não o contrário.

A autora assume uma estrutura subjacente para os nominais cujo nó que se deriva da sintaxe combina duas unidades mínimas: uma raiz – que identifica e distingue cada nome enquanto unidade do léxico – e um morfema abstrato que confere às

<sup>11</sup>Alcântara traz o pressuposto da não marcação do feminino fazendo referência a Câmara Jr., para o português, e Harris (1991) para o espanhol.

<sup>12</sup> *Redundancy Rules*



raízes suas categorias morfossintáticas (N, A, Adv etc.), tornando-as estruturas morfológicamente bem-formadas.

Ao fazer isso, Alcântara refere uma correlação inerente do gênero feminino com a terminação *a*, prevista por regra, enquanto que a correlação dessa terminação com o masculino não é regular, sendo necessária uma informação que a especifique. 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, a autora estipula uma regra em seu modelo que faz necessário que itens raríssimos em português como *tribo* carreguem a informação específica quanto ao segmento terminal *o*, de modo a impedir a aplicação da regra que relaciona itens femininos à vogal átona final *a*.

Armelin (2014) distancia-se da perspectiva modular a respeito das classes nominais e vai em direção a um tratamento unificado, 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.

Schwindt (2012, 2018) adota uma perspectiva otimalista realizacional, que une o Serialismo Harmônico (MCCARTHY, 2007) e abordagem otimalista Optimal Interleaving (WOLF, 2008) para dar conta dos nomes em português. Este modelo de análise, em certa perspectiva, distancia-se ainda mais do tratamento dado por Harris e Alcântara, que consideram diferentes módulos para as palavras da língua. Para o autor, os morfemas de gênero e classe temática são unidades independentes — por mais que fonologicamente idênticas em alguns casos —, exponenciando-se nos nomes com base em uma hierarquia realizacional em que diferentes traços competem, avaliados a partir de restrições universais, no curso de sua superficialização. Embora a análise não esteja limitada a apenas estes casos, seu foco recai sobre os nomes temáticos (fechados por *o*, *a* e *e*). 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; e CL2 e MASC), sendo, portanto, unidades diferentes do ponto de vista morfológico.

Para melhor entendermos a abordagem realizacional proposta por Schwindt (2011, 2018), observemos a formalização da hierarquia em (1).

(1)

(CL3), /e/ >> (FEM), /a/ >> (CL2), /o/ >> (CL1), /a/ >> (MASC), /o/

Nela, CL3 aparece como o morfema mais marcado, o que está de acordo com os dados de produtividade levantados no estudo: 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 “incomuns” 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*, embasa o ordenamento dos morfemas de classe temática CL2 e CL1. Por fim, o morfema MASC, substanciado por /o/, é tratado com o morfema menos produtivo na língua, uma vez que casos de pares sexuados como *menino/menina* dizem respeito a um /o/ temático na forma feminina, e não a um morfema de gênero, sendo agramaticais casos em que MASC seria expresso fonologicamente: *a professora/o \*professoro*. Dão conta, por outro lado, de formações encontradas na fala infantil, como *abelho*, *formigo*, *crianço* etc. Nessa fase de aquisição, a restrição que realiza *-o* como marca de gênero poderia estar mais alta na hierarquia do que no estágio estável da língua, em que tais formações são rejeitadas.

Dadas essas diferentes visões que nos permitem entender a organização do sistema de gênero e a forma como se distribui no léxico nominal da língua, prosseguiremos, na próxima subseção, para algumas abordagens quanto à localidade e o papel do gênero gramatical na sintaxe do português.

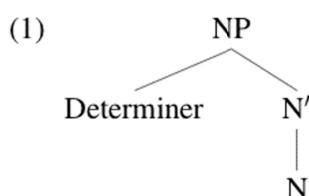
### 2.1.2.2.1 O gênero como uma categoria sintática

Nesta seção, apresentamos trabalhos que justificam a concepção de gênero enquanto componente da estrutura sintática dos nominais, ao lado, por exemplo, da categoria de número, em português. Ainda, trazemos propostas para caracterização desse traço morfossintático, sob o escopo da MD, que levam em consideração toda sua complexidade, que vem sendo demonstrada até aqui nesta seção, em especial pelo trabalho de Alcântara (2010).

Entendemos que, como ponto de partida para uma análise dos processos de atribuição de gênero a substantivos novos, nosso tema central, deve-se conceber uma proposta sobre a forma com que se estruturam os componentes internos dos nomes em português e se constitui o traço de gênero nessas estruturas. Explicar um fenômeno em que, p.ex., um falante de PB atribui o *feminino* a uma pseudopalavra exige que consideremos alguns aspectos do processamento e da produção linguística. Ao apresentar e propor aqui uma análise sobre um dado linguístico, nos interessam, por exemplo, pressupostos quanto à localidade do traço de gênero na estrutura do enunciado em questão (ou, e doravante, *locus* de gênero) e às etapas da derivação em que informações desse traço podem ser acessadas. Assim, os trabalhos discutidos a seguir fornecem bases importantes para a fundamentação teórica de nossas discussões em seções subsequentes.

É importante, primeiramente, compreender as implicações do modo como se concebe a estrutura nominal das línguas de gênero e a localização do traço em meio a ela. Inicialmente, conforme proposto por Jackendoff (1977), entendia-se a estrutura dos nominais como tendo em sua projeção máxima um núcleo SN:

Figura 3 – Estrutura com SN como projeção máxima.

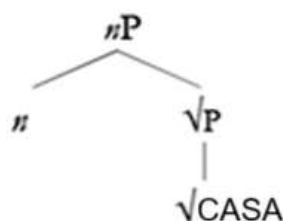


Fonte: Jackendoff (1977).

Nessa concepção, o *locus* das informações de gênero seria *N* (*Noun*: núcleos nominais, em português, substantivos e pronomes), o que implica papéis dos componentes de SN na derivação: *Det* não traz um traço de gênero enquanto unidade lexical, mas expressa um valor de gênero que é determinado por *N*.

Com início em Chomsky (1995)<sup>13</sup>, propõe-se que núcleos sintagmáticos como SN e SV compreendem elementos lexicais, como raízes nominais e verbais — então já assumidos nesse *locus* pela literatura —, mas também elementos funcionais, que representam uma série de propriedades, entre elas traço categorial<sup>14</sup>, caso, número, em uma hipótese que viria a ser chamada de decomposição lexical. Em um *NP* (SN) como *casa*, por exemplo (Figura 4) um elemento funcional *n*, forneceria o estatuto de nome — e a potencialidade de receber informações de número e gênero, por exemplo — a uma raiz  $\sqrt{\text{CASA}}$ . A hipótese redistribui informações anteriormente agrupadas nos núcleos (p. ex., *N* e *V*), as quais passam ao estatuto de raiz não categorizada, representando agora, enquanto unidades do léxico, suas identidades idiossincráticas. Aos núcleos funcionais, a *hipótese dos núcleos funcionais*<sup>15</sup> atribui os traços que subjazem processos gramaticais desde a derivação sintática. Desse modo, passam-se a conceber sintagmas cuja estrutura interna implica necessariamente a fusão de núcleos interdependentes.

Figura 4 – Estrutura de um SN sob a proposta da Decomposição Lexical.



Fonte: Jackendoff (1977).

Posteriormente à concepção de SVs como projeções de núcleos funcionais, passa-se a conceber a organização sintática de nominais também baseada em um elemento dessa natureza. Como representado na Figura 5 a seguir, pressupõem-se

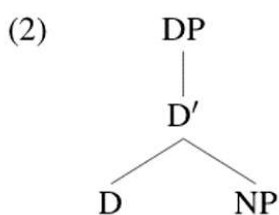
<sup>13</sup> Proposta difundida por proponentes da MD em trabalhos como o de Kratzer (1996), Harley e Noyer (1999) e Embick e Noyer (2001).

<sup>14</sup> Tradução nossa de *word-class*. O termo equivalente, empregado por muitos dos trabalhos referenciados nesta subseção é *category*.

<sup>15</sup> Do inglês *lexical head hypothesis*, conforme Grimshaw (1991), Alexiadou (2004) e Kramer (2015)

projeção máxima da estrutura nominal, *DP*, um sintagma determinante (em inglês, *DP*, de *determiner phrase*; aqui doravante, *SD*). O *locus* de *SN* passa a ser amplamente concebido (por HARRIS, 1991; ALEXIADOU, 2004; ALCÂNTARA, 2010, entre outros) em associação a um elemento funcional com base em análises do traço nominal de número.

Figura 5 – Estrutura nominal sob um núcleo *SD*.



Fonte: Alexiadou (2004)

Assim como os SVs e os traços verbais de caso, tempo, modo e aspecto, por exemplo, a categoria de número nominal dispõe de propriedades que possibilitam o tratamento como núcleo funcional. De acordo com Chomsky (1995), a Gramática Universal dispõe de uma série de traços, diferentes conjuntos dos quais são selecionados por cada língua. E neste sentido, assume que duas propriedades devem ser considerados ao se assumirem fenômenos linguísticos enquanto projeções sintáticas: i) a associação com a interface semântica, o que se trata pelo conceito de *interpretabilidade*; ii) os efeitos sobre a sintaxe, como participar de processos de movimento ou concordância. O traço nominal de número i) expressa na morfologia a quantidade de referentes denotados pelo nome em seu *SN* e ii) se associa na derivação ao processo de concordância e a operações de movimento de constituintes.

Importa destacar, por fim, que a realização morfológica de um traço não implica necessariamente relevância semântica ou sintática: o traço de caso dos pronomes em inglês, por exemplo, não se expressa morfológicamente, ainda que se associe à sintaxe e semântica.

No caso de gênero, segundo Alexiadou (2004), o único indicador que o fundamentaria é o fator concordância, não havendo, para a grande maioria das línguas de gênero, uma interpretação semântica regular ou uma distribuição morfossintática uniforme que o torne predizível em um dado nominal. Isso se deve

ao fato, de que, por exemplo, nem todos os nomes em português têm um gênero que se associa ao sexo do referente, com e de que as marcas relacionadas a gênero, embora existentes, não se instanciam em todos os nomes, como é o caso dos substantivos terminados por *e*.

A isso está associado ao fato de que as vogais finais de algumas línguas românicas, conforme apontado por Harris (1991) e Alexiadou (2004), não correspondem todas a marcas de flexão de gênero, mas sim a morfemas de classes flexionais, ou vogais temáticas. Isso se explica pela grande disparidade entre vogais terminais e outros segmentos, de um lado, e traços de gênero, de outro.

Assim, defende-se (ALEXIADOU, 2004; KRAMER, 2015, 2016) que gênero não possui sua própria projeção sintagmática na estrutura dos nomes, o que seria algo como GenP (*Gender phrase* ou sintagma de gênero), nem que o traço para essa categoria esteja localizado sob Num, o núcleo de número. Ao invés disso, costuma-se conceber a estrutura nominal como a localidade sintática de gênero, seja em N, ou, em abordagens que decompõem as categorias lexicais, no núcleo nominalizador *n*.

Aqui, mais uma vez, vem à tona a intersecção entre gênero e classe temática nos nomes do português. Alexiadou (2004) discute a localidade dos traços de gênero e classe<sup>16</sup> na estrutura sintática, partindo da relevância desses elementos em processos do domínio nominal. A autora discute a existência de um papel funcional na projeção estendida dos nomes, defendendo que esses traços não atuam sintaticamente no processamento. O traço de classe flexional das raízes nominais seria inerentes a sua estrutura, ao passo que gênero, por outro lado, pode também ser atribuído por meio de regras morfológicas, como a afixação de um sufixo portador de gênero (*mulher f.* > *mulherão m.*).

Quanto à relevância para o sistema sintático, Alexiadou assume para o traço de gênero um papel no processamento graças a sua atuação na concordância nominal. Enquanto que a caracterização de classe flexionais toma como base padrões da expressão morfológica de grupos nomes ou verbos — em português, por exemplo, *mesa*, *amiga* e *tema* compõem a classe dos nomes em *a*; e os verbos *sair*, *dormir* e *cair* apresentam o mesmo morfema de classe *i*, que precede os demais afixos, como o *-r* infinitivo nestes exemplos —, o pertencimento de um dado

<sup>16</sup> Traduzimos aqui *inflectional class* por classe flexional, ou simplesmente classe.

nome a uma classe de gênero está associado à concordância dos elementos coordenados de seu SN. Assim, artigos indefinidos compondo o sintagma nominal de *casa*, *torre* ou *pá*, todos femininos, mas de classes flexionais distintas, terão a mesma forma flexionada *uma*, em português.

Em seu trabalho, que se dedica à caracterização sintática dos sintagmas nominais, Alexiadou constata o fenômeno que é objeto central deste estudo ao abordar o “caráter misto” do traço de gênero, que se explicaria pela natureza da sua associação com os radicais nominais. Nessa perspectiva, a autora opõe casos de gênero “fixo”, especificado no léxico, a casos de gênero “variável”, não inerente, mas atribuído a radicais marcados como [+humano/+animado] por meio de “concordância” entre propriedades semânticas do referente e classes de gênero. A relevância desse fato para a discussão em Alexiadou parece estar relacionada ao comportamento característico dos nomes de “gênero variável”, atestado em todas as línguas consideradas pela autora.

A atribuição de gênero aos substantivos explica-se como propriedade inerente do radical nominal, para os substantivos inanimados — uniformes, portanto —, e como derivada de propriedades do referente, no caso de substantivos humanos e animados, de gênero variável. No segundo caso, a autora assume que deve ocorrer uma relação entre um nominal e seu referente humano, cujos traços relevantes são então copiados para o nome: um referente masculino animado como *gato*, projetaria o traço semântico de sexo sobre a morfossintaxe, onde corresponderia ao gênero gramatical masculino.

Em outro trabalho sobre a sintaxe do gênero gramatical, Kramer (2016) aponta para uma diferença crucial para a discussão da localidade sintática do gênero gramatical, o fato de esse traço poder estar relacionado tanto a um marcador de concordância quanto à atribuição de gênero em si. Para a abordagem,

Para a análise que vê a localidade em N, o traço de gênero faz parte da raiz, sendo parte de suas informações idiossincráticas.

Algumas abordagens, normalmente ligadas à MD, entendem cada categoria lexical como sendo decomposta em raiz e categorizador, o qual faz com que raízes abstratas e sem especificação de classe gramatical passem a pertencer a uma dessas classes. Nessa visão, o *locus* de gênero em geral é *n*, o elemento nominalizador que se combina às raízes para formar núcleos nominais *nP*. Um dos

ganhos dessa visão é a possibilidade de se tratarem raízes de nomes como *dentista* como desprovidas de gênero, tendo feminino e masculino atribuídos a depender do categorizador a que se combinam. De outro modo, nessa perspectiva teórica, seriam necessárias duas raízes para *dentista* com mesma configuração fônica e com significados quase idênticos. Além disso, o gênero gramatical atua em processos relacionados a *n*, tais como a nominalização.

Observamos que as abordagens que associam gênero a um nó nominalizador, como as que estamos vendo aqui desde a referência Alcântara (2010) o fazem em princípio por uma questão de economia para o modelo teórico. A informação de gênero entendida como contida em um sufixo independente, mas é nesse sentido vantajoso entendê-lo como parte de um nó categorizador que é necessário para a boa formação morfológica ao lado das raízes. Ainda, entender gênero como atrelado ao nominalizador condiz com o fato de ser esta uma propriedade exclusiva dos nomes.

Em uma abordagem decomposicional sobre o léxico, um sistema como o português, com gêneros feminino e masculino, teria dois nominalizadores *n*, um marcado para gênero e outro não, sendo que condições de licenciamento combinam as raízes dos substantivos a um ou outro desses categorizadores.

Essa subcategorização também ocorre para o caso do mapeamento semântico de traços do referente para a matriz morfossintática dos nomes. Quanto a propriedades semânticas dos substantivos que estão relacionadas ao gênero, como o fato de femininos denotarem entidades do sexo feminino, as abordagens que concebem gênero em N precisam lançar mão de regras lexicais que registrem na raiz uma especificação como 'female biological sex' (HARRIS, 1991), o que atribui a elas o gênero gramatical feminino.

Uma vez que, por outro lado, as abordagens decompositivas normalmente não admitem um grande número de informações no léxico propriamente dito, é o categorizador *n* o responsável pela interface do gênero gramatical com a semântica. Para Kramer (2015), isso se traduz na identidade do traço de gênero, que pode ser ou não interpretável do ponto de vista semântico. Nessas perspectivas, traços sintáticos variam em sua interpretabilidade<sup>17</sup>. Assim, raízes de

---

<sup>17</sup> A interpretabilidade dos traços diz respeito à leitura de uma informação sintática no componente semântico, por exemplo o traço de número, responsável pela expressão morfológica do substantivo e a flexão dos alvos da concordância, também indica uma categorização conceitual do referente.



nominais animados combinam-se a raízes interpretáveis e respectivamente para os nominais inanimados. Animados como *persona* ou *membro* só podem ser combinados a *n*'s não interpretáveis (KRAMER, 2015).

(2) Classificação dos núcleos nominalizadores do espanhol

a)  $n_u[-FEM]$  – núcleo com feminino não interpretável

ex: *ouro, problema, galpão*

b)  $n_u[+FEM]$

ex: *rede, classe, conclusão*

c)  $n_i[+FEM]$

ex: *candidata, dona, moça*

c)  $n_i[-FEM]$

ex: *candidato, dono, moço*

Kramer elenca três motivos pelos quais entende uma leitura de gênero localizado em *n*, por uma abordagem de decomposição lexical, como mais adequada. Primeiro, ela evoca a hipótese Borer-Chomks, segundo a qual variação paramétrica não pode estar relacionada a cabeças lexicais, mas apenas a cabeças funcionais. Uma vez que há variação paramétrica de gênero (*sangre* e *sangue*, por exemplo), não poderia gênero estar alocado em N, mas não haveria problemas em estar em *n*, um nó funcional.

Em segundo lugar, a autora aponta a distinção entre gênero gramatical e sexo biológico<sup>18</sup> atentada por análises como a de Harris (1991) pode ser associada à diferença entre (2.a–b) e (2.c–d). Em uma abordagem que entende que gênero se atribui a *n*, feminino e masculino residem nesse mesmo traço, que pode ou não ser interpretável semanticamente.

Por último, a autora traz dados do amárico, em que alguns substantivos com referentes sexuados apresentam o feminino como *default*, expresso em casos em que não se sabe o sexo do referente, mesmo sendo o masculino o *default* da língua como um todo. Kramer (2015) explica o fato pela possibilidade combinatória da raiz *ayt* (rato), por exemplo, tanto com [-FEM] e [+FEM] interpretáveis, como com

---

<sup>18</sup> Entendemos aqui que *sexo biológico*, mas também *gênero social* são propriedades conceituais dos referentes animados, agrupando-se sob uma mesma categoria semântica.

[+FEM] não interpretável, mas nunca com [-FEM] não interpretável, uma vez que a única possibilidade da expressão de feminino por esta palavra seria diante de um referente fêmea (por interpretação semântica).

As discussões quanto à localidade do traço gênero na estrutura sintática dos nominais e bem e a natureza e classificação dos traços que informam quanto à categoria, por exemplo, no nó categorizador, serão retomadas ao fim da próxima seção, bem como nas discussões sobre nossas análises empíricas, para explicar a atribuição de gênero a novas formações.

## 2.2 O gênero gramatical em processos psicolinguísticos

Segundo Corbett (1991), o processo de atribuição de gênero gramatical poderia ser observado *in loco* em, pelo menos, três tipos de fenômenos linguísticos: a aquisição como primeira língua ou como língua adicional de gramáticas de gênero, o processamento de pseudopalavras em experimentos psicolinguísticos e a formação de empréstimos nominais. Segundo o autor, tais fenômenos compõem o que entende como a *realidade psicológica*<sup>19</sup> dos sistemas de atribuição de gênero gramatical, e serão abordados individualmente nos itens a seguir.

### 2.2.1 Aquisição de sistemas de gênero

Desde seus primeiros anos de vida, falantes de línguas naturais aprendem, de forma espontânea, uma série de estruturas e processos complexos que fazem parte da língua, ou línguas, falada/s em seu meio. Segundo Lightbown e Spada (2020), se podem observar certos padrões universais no modo e na ordem com que crianças desenvolvem sua competência linguística. Os substantivos, em geral a classe das primeiras palavras a serem aprendidas, antecedem a aquisição de verbos, os quais tendem a ser usados antes dos advérbios, por exemplo. Desde os primeiros estágios de aquisição, se pode notar certa regularidade no uso de morfemas. Em inglês, a aquisição do sufixo do presente contínuo (*-ing*) ocorre antes dos morfemas regulares do passado simples (*-ed*), nos verbos, ao passo que, nos nomes, o *-s* plural aparece antes dos artigos (LIGHTBOWN; SPADA, 2020, p. 24).

A aquisição de novas formações lexicais é um processo inerente não só aos estágios iniciais do desenvolvimento linguístico de crianças enquanto adquirem

---

<sup>19</sup> *The psychological reality of the gender assignment systems (Idem, p. 70)*

suas primeiras línguas, mas também recorrente na ampliação do léxico de falantes adultos. Também de acordo com Lightbown e Spada, em *How Languages are Learned* (2020), a aquisição de novas palavras pode ocorrer tanto de forma intencional, por aprendizado com instruções explícitas, quanto de forma *incidental*<sup>20</sup>. Na aquisição incidental, segundo as autoras acontece “*leitores e ouvintes fazem uso de fontes variadas de informação, incluindo o contexto, a sintaxe e a morfologia, para inferir o significado de palavras que não conhecem*” (LIGHTBOWN & SPADA, 2020, p. 146)<sup>21</sup>.

Especificamente sobre a aquisição de sistemas de gênero gramatical, Levy (1988) afirma, referindo estudos anteriores sobre aquisição, que aos três anos crianças monolíngues já seriam capazes de processar a categoria de gênero. Nesse período da aquisição, segundo o autor, em geral já não se atestariam erros de concordância, por exemplo, sendo que os poucos casos estariam relacionados a substantivos novos ou a traços extralinguísticos associados a gênero. Aos quatro anos de idade, no entanto, mesmo esses erros tenderiam a desaparecer por completo.

Nesta seção, apresentamos trabalhos dedicados à aquisição de substantivos, em vista da atribuição de seu gênero gramatical, seja por casos de crianças aprendizes de língua nativa ou de adultos no uso de línguas adicionais. A aquisição de sistemas de gênero de primeiras línguas aparece como objeto de trabalhos em maioria recentes, sendo que, da publicação desta dissertação, não encontramos investigações sobre o fenômeno com foco em variedades do português. Assim, apresentamos na sequência estudos do espanhol (PÉREZ-PEREIRA, 1991), do neerlandês (BLOM; POLISENSKÁ; WEEMAN, 2008), do grego moderno (TSIMPLI; HULK, 2013; KRAAIKAMP, 2017) e do isangu (IORIO, 2017).

Como ponto de partida para nossas discussões sobre aquisição de sistemas de gênero, apresentamos em (4) a seguir os *princípios de complexidade* propostos por Audring (2014). A autora entende *complexidade* como um parâmetro que pode ser aplicado a diferentes aspectos de um sistema de gênero, tais como o número de valores de gênero, as propriedades morfossintáticas dos marcadores de gênero e a atribuição semântica de substantivos em relação aos valores de gênero. A

<sup>20</sup> No original, *incidental vocabulary acquisition* (*Idem*, p. 146).

<sup>21</sup> Tradução nossa de: *Readers and listeners use various sources of information, including context, syntax, and morphology, to infer the meanings of unfamiliar words.*

avaliação da complexidade de um sistema de gênero estaria diretamente implicada a sua aquisição, uma vez que poderia apontar fatores específicos que dificultam seu processamento.

(4) Princípios de complexidade de gênero

- a) Princípio de Economia: a complexidade de um sistema corresponde ao número de classes de gênero que expressa;
- b) Princípio de Transparência: um sistema minimamente complexo apresenta mapeamento 1:1 entre gênero e forma;
- c) Princípio de Independência: a complexidade é mensurada pela associação de gênero e suas marcas com propriedades de outras dimensões linguísticas ou extralinguísticas. (adaptado de AUDRING, 2014, p. 5–6)

Tomando-se o caso do PB para demonstrar a aplicação da metodologia de Audring, temos que: embora se ateste o uso de outras expressões, de modo geral, há dois valores que podem ser expressos pela língua: *masculino* e *feminino*, sendo assim um sistema minimamente complexo em relação a (4.a). Como veremos na sequência, línguas como isangu, o russo e o grego, por sua vez, podem ser mais complexas quanto a esse princípio, apresentando três ou mais valores de gênero.

Quanto a (4.b), os nomes do português, como vimos nas diversas descrições em 2.1, não são totalmente transparentes quanto a gênero, em especial os substantivos. Ainda assim, há associações regulares, como do *feminino* e a terminação *a*. No entanto, adjetivos, pronomes e principalmente artigos se aproximam do mapeamento 1:1 idealizado em (4.b): *pequena* e *grata*; *ela* e *aquela*; e *a* e *uma*, respectivamente. Audring (*Idem*, p. 5, citando BENTLEY & KULEMEKA, 2001) traz como exemplo de transparência nesse sentido a língua chichewa, cujo valor de gênero 7/8<sup>22</sup> é expresso pelos substantivos e por todos os elementos concordantes pelas mesmas marcas *chi-/zi-*, com a exceção de um pronome demonstrativo.

O *Princípio da Independência* em (4.c) é de especial interesse às nossas análises e discussões. Embora esse princípio afirme que um maior número de associações entre gênero e propriedades externas a ele agrega complexidade a

<sup>22</sup> Como veremos também no estudo apresentado ao fim deste item 2.2.1, os valores de gênero gramatical das línguas bantu são convencionalmente assim representados, correspondendo a pares de classes plural e singular. Nesse caso, 7 é o valor de gênero expresso pelo plural, e 8, para a forma singular.

seu sistema, assumimos aqui que é a existência dessas associações que explica a atribuição de gênero a novas formações. O que para Audring corresponderia à noção de “dependência” diz respeito à referência a propriedades externas ao sistema de gênero propriamente dito, as quais devem fundamentar nossas análises dos dados analisados nos Capítulos 3 e 4 e são consideradas nos estudos de caso a seguir.

Pérez-Pereira (1991) conduziu um experimento psicolinguístico para investigar a atribuição de gênero em espanhol por falantes em diferentes etapas da aquisição. O estudo contou com a participação de 160 crianças nativas da língua, oriundas das regiões de Santiago e Vigo, sendo metade meninas e metade meninos, pertencentes a diferentes faixas etárias, entre quatro e onze anos, e das classes média e média-baixa.

Com o objetivo de investigar os fatores que influenciam o processo de atribuição de gênero, o autor testou três tipos de regras: sintáticas, morfofonológicas e extralinguísticas. Para avaliar o papel individual de cada tipo de regra, foram utilizadas pseudopalavras que continham de um a três desses índices, bem como suas possíveis combinações.

As pistas sintáticas eram mostradas como determinantes flexionados que vinham antes dos substantivos-teste. Já as pistas morfofonológicas estavam presentes nas terminações das palavras — como *a* para o feminino e *o* para o masculino. As pistas extralinguísticas eram apresentadas por meio de imagens — desenhos que representavam humanos com características estereotipicamente associadas ao feminino ou masculino — expostos ao lado das palavras-teste.

Pérez-Pereira conclui sugerindo que pistas linguísticas têm um papel preponderante na atribuição de gênero, se sobrepondo a informações extralinguísticas. Ou seja, ao categorizar os pseudosubstantivos como masculinos ou femininos, independentemente de fatores sociais ou faixa etária, os falantes levariam em conta principalmente a forma dos substantivos ou de determinantes associados.

Por fim, quanto à variável faixa etária, o autor afirma que, nas fases iniciais da aquisição de gênero, que vão dos quatro aos seis anos de idade, os falantes apresentam uma preferência em relação aos índices morfofonológicos — terminações associadas a gênero —, em comparação aos sintáticos — determinantes flexionados. Segundo Pérez-Pereira, essas pistas morfofonológicas

parecem estar disponíveis mesmo antes da aquisição completa da concordância nominal, que ocorreria por volta dos cinco anos. A partir dos seis anos, no entanto, devido a essa aquisição completa da sintaxe da concordância, os participantes passariam a basear suas escolhas mais na dimensão morfossintática e menos na morfofonológica, ou seja, diante de contextos sintáticos que incluem artigos definidos precedendo as pseudopalavras, a atribuição do gênero esperado ocorreria na maioria dos casos.

Em Blom, Polisenská e Weerman (2008), é abordada a aquisição do neerlandês por três grupos distintos: crianças monolíngues, crianças bilíngues e adultos aprendendo o idioma como língua adicional. O neerlandês é uma língua que possui três gêneros gramaticais, o masculino, o feminino e o neutro. Os autores observam que, de forma bastante recorrente, os três grupos atribuem o gênero incorretamente aos substantivos nos primeiros estágios de aquisição. O fenômeno em que o gênero masculino é atribuído a formas femininas ou neutras é, segundo o estudo, um processo de hipergeneralização do artigo masculino a nomes neutros. Por outro lado, a análise aponta que, enquanto crianças operam a hipergeneralização de um sufixo adjetival em particular, os adultos usam tanto formas sufixadas como não sufixadas incorretamente. A diferença estaria relacionada ao fato de que adultos tendem a se basear no *input*, seguindo um caminho de aquisição orientado pelo léxico, enquanto crianças fazem uso de representações baseadas na gramática. Ou seja, segundo os dados estudados, falantes apresentados a um volume lexical maior, com mais casos excepcionais onde terminações são atipicamente associadas a um gênero, interfeririam na atuação de uma regra mais simples que atribui sufixos aos gêneros de maneira mais objetiva.

Em relação ao grego moderno, língua que também possui um sistema tripartite de gênero, com feminino masculino e neutro, Tsimpli e Hulk (2013) assumem que o gênero neutro é o *output default* dos processos de atribuição. Ainda assim, notam que a manifestação do neutro pode estar ligada a dois casos diferentes: i) a avaliação de traços relacionados com o neutro durante a computação do nome e ii) a falta de pistas relevantes durante a computação, quando o neutro se superficializa por *default*. Desse ponto de vista, os autores estipulam duas classes distintas sob o que tradicionalmente se considera como

neutro, uma delas sendo a não marcada e a outra contando com um esquema de marcação parecido com os outros dois gêneros da língua, feminino e masculino.

Embora no grego esses dois casos não possam ser observados em contraste, há casos similares quanto ao fenômeno de atribuição de gênero em neerlandês (RODENBURG, HULK; 2008; 2010), quando se afirma que o neutro pode ser assumido como *default* em alguns contextos. Assim, mesmo admitindo ser o neutro a forma não marcada na atribuição durante a aquisição, Tsimpli e Hulk (2013) notam um período instável onde o artigo de gênero comum *de* (relacionado a masculino e feminino) também emerge como *default*.

De novo, o pressuposto dos autores quanto a um gênero adicional, não *default*, que se deriva da classe de gênero normalmente assumida como *default* é respaldado pelos dados do neerlandês mencionados acima. Nos casos em que neutro e comum parecem competir para a expressão como *default*, a presença de um sufixo diminutivo *-(t)je*, por exemplo, o qual está estatisticamente associado ao neutro no léxico da língua, aparece como um critério decisivo para a atribuição do neutro, sendo computado invariavelmente para cada classe.

Finalmente, e ainda quanto à instabilidade de um *default* para a atribuição, Tsimpli and Hulk (2013) coloca o fato como uma pista para processos de mudança do sistema de gênero gramatical do grego. É que, uma vez que os pronomes da língua associam explicitamente classes de gênero e traços semânticos — comum para animados, neutro para inanimados — esse sistema, pronominal, é processado consistentemente pelas crianças monolíngues na aquisição (como é o caso do neerlandês). Por outro lado, os autores defendem que o retardo na aquisição da atribuição de gênero do grego de forma consistente, evidenciada pela instabilidade na escolha de uma classe *default*, está associada como uma perda ou empobrecimento em transe do sistema de gênero gramatical.

Partindo para a última análise, Iorio (2017) apresenta dados da língua isangu em sua aquisição como primeira língua. Notam-se peculiaridades importantes no modo com que crianças em diferentes fases da aquisição atribuem gênero a certos substantivos. Os principais desvios envolvem a generalização da marcação de um dos tipos de gênero sobre raízes pertencendo aos demais gêneros. Dentre os fenômenos mais recorrentes, destacamos: i) a atribuição do gênero denominado

1/2<sup>23</sup> raízes nominais que, no léxico adulto, pertencem à classe 2/9 — pela afixação de *ba-*, ao invés de *tsi-*; e ii) a atribuição generalizada do gênero ½ a substantivos com referentes animados, i.e., o fato de sua atribuição a inúmeros substantivos que, embora pertençam a outros gêneros, denotam referentes animados.

Assim, vimos que o gênero gramatical é uma das propriedades gramaticais que se adquirem pelos falantes, sejam eles crianças aprendizes de primeira língua, ou adultos em seu dia a dia, a quem são apresentados novos substantivos, demaneira incidental ou por instrução explícito. Em todos os casos, os mecanismos envolvidos no processamento de gênero compartilham de alguns aspectos. Uma das maneiras de controlar alguns desses aspectos e verificar em maior profundidade o fenômeno é através de experimentos psicolinguísticos.

### 2.2.2 Experimentos psicolinguísticos

Clegg (2010) apresenta um experimento com pseudopalavras para investigar como falantes nativos de espanhol atribuem gênero aos substantivos em na língua. Seu estudo teve como objetivo determinar se há regras ou padrões específicos que os falantes de espanhol usam para atribuir gênero aos substantivos, questionando a visão de que o processo se opera por fatores subjetivos ou arbitrários<sup>24</sup>.

A partir de levantamentos de outros estudos e do léxico dicionarizado, o autor elencou uma série de sufixos nominais que estariam estatisticamente associados ao gênero feminino, como *-dad* e *-umbre* (de palavras femininas como palavras *claridad*, claridade e *costumbre*, costume), os quais considera fundamentais para a atribuição de feminino nas pseudopalavras, como *alaridad* e *colidumbre*, criadas especificamente para o estudo.

Clegg conduz um experimento que testa o processamento desses sufixos para o gênero feminino através das respostas dos falantes às questões. Durante a realização do teste, os participantes eram apresentados a uma lista de pseudopalavras, compostas por um sufixo associado ao feminino precedido por segmentos artificiais, e eram então solicitados a atribuir um gênero a cada item.

Os resultados mostraram que os falantes nativos de espanhol foram capazes de atribuir o gênero feminino às pseudopalavras com os sufixos-teste com um alto

<sup>23</sup> Como já mencionado na página 39 da seção 2.1, é assim que se representam os valores de gênero das várias línguas bantu.

<sup>24</sup> Alguns dos primeiros estudos a abordarem o tema da atribuição de gênero a substantivos chegavam a conclusões como essa.



grau de precisão, indicando que existem regras ou padrões específicos que regem a atribuição de gênero em espanhol.

O estudo também descobriu que houve algumas exceções a essas regras, e que a atribuição de gênero nem sempre foi consistente entre os participantes. Isso sugere que, embora existam padrões gerais que regem a atribuição de gênero em espanhol, também há algum grau de subjetividade ou variação na forma como os indivíduos atribuem gênero a palavras específicas.

Em suma, o estudo fornece *insights* sobre como os falantes nativos de espanhol intuitivamente atribuem gênero a substantivos novos em sua língua com base em associações entre constituintes formais e padrões do léxico internalizado. Ainda, o instrumento utilizado pelo autor para testar sufixos nominais enquanto índices formais de atribuição de gênero em espanhol serviu como referência para um dos experimentos que apresentamos em trabalhos anteriores, e aos quais a presente dissertação dá seguimento.

Em Surreaux (2020), apresentamos dois experimentos que têm como objetivo testar os critérios formais — fonológicos, grafêmicos e morfológicos — e semânticos envolvidos na atribuição de gênero a pseudopalavras em português. No teste de índices formais, foram utilizadas pseudopalavras elaboradas a partir de segmentos finais recorrentes que estão estatisticamente relacionados ao gênero feminino, como os sufixos *-ção*, *-agem*, *-ade*, *-ite*, *-ose* e o segmento *e*. Os 89 participantes do experimento eram falantes nativos de português e responderam a um questionário que trazia frases com as pseudopalavras. Cada frase possuía uma lacuna que deveria ser preenchida com um dos dois artigos definidos do português, tornando explícito o gênero que cada sujeito atribui aos substantivos.

Os resultados indicaram que a maior parte das terminações testadas coincidiram com a atribuição de feminino, as quais tiveram esse gênero atribuído com frequências de 88,5% a 100%.

Um segundo experimento do mesmo estudo investigou a atribuição de gênero gramatical com base em critérios semânticos, também utilizando pseudopalavras. Dessa vez, foram elaboradas de modo que não fornecessem índices formais relacionados a gênero, de modo a isolar a motivação semântica. As pistas testadas eram apresentadas juntamente a cada pseudopalavra na forma de breves definições do que seriam os seus referentes, que poderiam ser animados ou inanimados, como nos exemplos em (6):

- (6) a) Horb: um tipo de cadeira de balanço.  
b) Melateb: o mais novo dentre os irmãos do rei.

Assim, assumimos que o gênero de *horb* (6.a) seria resgatado do substantivo feminino *cadeira*, já presente no inventário lexical dos participantes, do mesmo modo que a informação de gênero de irmão(s) em PB levaria à atribuição do masculino a *melateb* (6.b). Esses casos exemplificam um processo que será referido aqui como atribuição de gênero por analogia, tomando como base a abordagem de Thornton (2009) sobre a atribuição de gênero gramatical a novos substantivos, exposta na sequência.

Os resultados demonstraram que o entendimento de um referente animado, de *sexo* ou *gênero social* feminino, associado a um novo substantivo parece ter maior influência que sobre a atribuição do feminino gramatical se comparado a informações sobre um referente inanimado, mesmo que relacionado a um nome feminino já conhecido, ou seja, os participantes consideraram que *Sark, a fêmea do rinoceronte* era uma palavra feminina em 90% dos casos, enquanto que *Zarf, ponte elevadiça que dá acesso a castelos ilhados*, foi processado como feminino em 74,3% das vezes.

### 2.2.3 O gênero em empréstimos nominais

Nesta seção, buscaremos delimitar o fenômeno da atribuição de gênero gramatical na adaptação morfológica de substantivos que são introduzidos como empréstimos. Para isso, serão abordados trabalhos da literatura que colaboram para a caracterização dos empréstimos enquanto fenômeno linguístico e para seu enquadramento no campo dos estudos linguísticos, sendo eles Carvalho (1986), Haspelmath (2009), Haugen (1950, 1956) e Poplack (2020). Além disso, serão abordados alguns trabalhos sobre o funcionamento da atribuição de gênero em casos específicos de empréstimos nominais em línguas com sistema de gênero nominal. Tais trabalhos incluem Barkin (1980), Poplack, Pousada e Sankoff (1982), Corbett (1991), Kilarski e Krynicki (2005) e Thornton (2009).

Haspelmath (2009) afirma que a formação de palavras novas por meio de empréstimos lexicais é um fenômeno intrínseco às sociedades humanas e a todas as línguas naturais. Sua afirmação incisiva é justificada pela prática migratória,

praticada pelos grupos humanos (DIAMOND; 1997). natureza humana. uma vez que é não pode ser dissociado do comportamento migratório dos humanos.

O contato linguístico é responsável por uma variedade de fenômenos interessantes aos estudos linguísticos. Processos decorrentes do contato entre línguas podem ser observados em nível fonológico, morfológico, sintático e semântico, estando normalmente associados à implementação de novos itens lexicais em uma língua a partir de outra (SANKOFF, 2001). Ao ser integrado ao léxico da língua receptora e estar sujeito a seus processos gramaticais, um empréstimo lexical necessariamente se adapta às estruturas do novo sistema (HAUGEN, 1956).

Conforme elencado em Carvalho (1986), o fenômeno dos empréstimos está entre os processos que possibilitam a ampliação lexical, ao lado da alteração de um item pré-existente; da seleção de uma variante em detrimento de outras em um sistema estável; da criação sistemática de novos itens a partir de regras internas ao sistema; e da economia funcional, em que distinções pouco salientes são apagadas. Notamos que o fenômeno aqui investigado diz respeito não só aos empréstimos nominais e sua acomodação ao sistema receptor, mas também ao caso análogo da ampliação lexical por meio de pseudopalavras em ambiente experimental, não considerado nas caracterizações da literatura, possivelmente por seu caráter artificial.

Uma vez que parte de nossa investigação se dedica à análise dos empréstimos lexicais, fazem-se necessárias algumas considerações a respeito desse fenômeno, em vista de sua caracterização. Ao entender os empréstimos lexicais como produto de processos de mudança linguística, Haspelmath (2009, p. 38) sugere que uma dada palavra em uma língua pode sempre ser um empréstimo em potencial, podendo ter sido implementada em estágios diacrônicos que escapam ao conhecimento científico. Valemo-nos dessa visão diacrônica sobre os empréstimos para atentar ao recorte a ser abordado em 4.1. e 4.2. Por mais remota que seja sua implementação, e ainda que entendidas por parte dos falantes como parte do léxico nativo do português, o *status* das palavras de origem árabe e tupi enquanto empréstimos, somado ao pressuposto uniformitarista quanto à mudança linguística, justifica as análises e discussões aqui propostas quanto à atribuição de gênero a itens novos em português.

Haspelmath (2009) entende empréstimo lexical (*loanword*) como uma palavra que passa a fazer parte de uma língua em algum ponto de sua história a partir de outra língua. Segundo o autor, as formas de origem dos empréstimos podem ser complexas ou mesmo frasais, mas perdem sua estrutura interna ao serem implementadas no novo sistema, onde são, em princípio, não analisáveis. Haugen (1956) traz mais considerações interessantes para a tipologia dos empréstimos lexicais e para a caracterização de seus processos na gramática da língua receptora: “Qualquer similaridade entre [as línguas envolvidas] é importação, enquanto qualquer diferença entre elas é vista como substituição do material nativo. Substituição quer dizer que a imitação do modelo estrangeiro é menos que perfeita, mas também quer dizer que passou a se tornar mais familiar àqueles que falam a língua nativa.” (HAUGEN, 1956, p. 388, tradução nossa)

A classificação dos empréstimos lexicais quanto à fidelidade às formas de origem se faz relevante por caracterizar os recortes de empréstimos apresentados na seção 4. Haugen (1956) divide os empréstimos lexicais em *loanwords*, em que tanto forma quanto significado são reproduzidos, como *tiquete*, *online*, etc.; *loanblends*, em que o empréstimo conserva significado e parte da forma — seria o caso da forma *goleiro* em português, em detrimento do empréstimo do equivalente *goalkeeper* —; *loan shifts*, onde apenas significado é importado — como *baixar* para o verbo do inglês *download*; e *loan translations*, em que é importado apenas significado, sendo a forma dada pela tradução literal da forma de origem, como os calques *cachorro-quente* e *palavra-chave* (para *hot dog* e *keyword*). São de nosso interesse os empréstimos lexicais caracterizados por Haugen como *loanwords*, uma vez que a integridade da forma de origem, vazia para o traço de gênero, garante o contexto para o fenômeno investigado.

Na adaptação de empréstimos, é possível que cada falante aplique um padrão individual ao novo item, propondo-o tacitamente através de seu uso contínuo à comunidade. É inapreensível a diversidade de fatores envolvidos em cada evento de produção e processamento dos empréstimos em seus estágios iniciais, ainda assim, itens lexicais emprestados gradualmente se estabilizam em novos sistemas linguísticos. Empréstimos lexicais são frequentemente classificados pelo seu grau de implementação ao sistema receptor. Comumente utiliza-se como

critério o grau de adaptação fonológica ou mesmo grafêmica dos itens para diferenciá-los de casos de *code switching* ou de estrangeirismos, por exemplo.

Poplack (2020) aponta para um período de inconsistência dessas novas palavras na gramática, onde os empréstimos apresentam variações fonológicas e morfológicas, por exemplo, sendo mais expressivas que aquelas de produtos léxicos de outras naturezas. É nesse período que há variação de gênero em substantivos essencialmente uniformes. O modo como se classificam palavras em relação a esses critérios, no entanto, como nota a autora, depende do objetivo da análise a ser feita sobre os itens. Neste trabalho, como veremos na subseção 4.3, dado o empréstimo em plena atividade de palavras do inglês no português brasileiro, consideramos como empréstimos, e não como casos de *code switching*, embora a introdução de palavras em ambos os casos compartilhem de certas particularidades e podem ser descritas pelos mesmos mecanismos teóricos, como veremos no final desta seção.

Poplack, Pousada e Sankoff (1982) é um dos primeiros estudos a abordar a questão da atribuição de gênero a empréstimos. Os autores utilizaram um corpus de 300 horas de entrevistas com falantes de espanhol porto-riquenho e francês de Montreal, de onde extraíram e compilaram todos os empréstimos do inglês detectados: um total de 872 ocorrências de 345 diferentes itens. Exemplos de segmentos contendo empréstimos incluem: “*Yo creo que mi hair está linda, beautiful*” (“Acho que minha *hair* está linda, *beautiful*”) e “*Tes hot dogs sont pas prêtes, la steam, elle vient pas*” (“Seus *hot dogs* não estão prontos, o *vapor* não está vindo”) (POPLACK; POUSADA; SANKOFF, 1982, p. 14, grifo nosso).

Os autores consideram alguns critérios para a categorização dos itens. O primeiro critério, de “sexo fisiológico” (*physiological sex*), diferencia os substantivos entre aqueles que possuem um referente animado, que por sua vez pode diferenciar entidades masculinas e femininas de acordo com o gênero gramatical, e aqueles que não apresentam tal potencialidade. O critério de “tradução analógica” (*analogical translate*), por sua vez, identifica uma expressão formal e semanticamente similar na língua receptora, p. ex., *establishment*, do inglês e *établissement*, do francês. Por último, considera-se a analogia com sufixos associados a um gênero na língua receptora, ou seja, se um empréstimo do inglês é fechado pela terminação *-tion*, é possível a analogia com o sufixo *-ción*, do espanhol, associado ao feminino.

Os achados do estudo apontam para a categoria semântica que distingue referentes de acordo com o “sexo fisiológico” (sexo biológico ou gênero social) como único critério significativamente regular para a atribuição de gênero. De modo geral, os critérios formais considerados no estudo variaram amplamente nas amostras analisadas, entre uma língua e outra e em cada uma delas tomadas separadamente.

Embora considerando a inexpressividade de critério fonológico para a atribuição de gênero, os autores chamaram atenção para o sufixo *-a* em espanhol — que, assim como no português, marca o gênero feminino — como apresentando relativa regularidade, algo não verificável no francês. Os autores ainda ressaltam a relativa instabilidade dos empréstimos quanto ao gênero gramatical, excetuando-se os casos em que há um referente animado que suporte relação isomórfica entre sexo biológico/gênero social e gênero gramatical. Os autores notam que os itens passam por períodos em que mais de um gênero é adotado para cada empréstimo na comunidade de fala. “Ouvimos *le job, une interview* em Paris, mas *la job, un interview* em Montreal” (POPLACK; POUSADA; SANKOFF, 1982, p. 26).

A discussão apresentada pelos autores é pertinente à nossa investigação, primeiro, por reforçar a ideia da regularidade da atribuição de gênero gramatical influenciada pela correlação com traços semânticos de gênero e sexo; segundo, por pontuar os estágios de latência dos empréstimos nominais, quando ainda não há adaptação morfológica completa, produzindo os casos de variação de gênero em substantivos uniformes. Ainda que substantivos emprestados venham a se tornar estáveis no gênero masculino, interessam os critérios para atribuição de feminino em seus estágios iniciais. Ainda, em análises futuras, visamos à descrição e à análise de empréstimos recentes sujeitos a esse fenômeno e os fatores que estariam envolvidos na acomodação a um ou outro gênero.

Corbett (1991) também aborda a questão da atribuição de gênero a novos itens, em seção dedicada à realidade psicolinguística do gênero gramatical. O autor propõe uma base explanatória para o fenômeno que serviu de referência para muitos estudos subsequentes — entre eles o de Thornton (2009), que se refere à obra, *Gender*, obra como “a bíblia dos estudos de gênero” (p. 34).

O autor classifica as línguas que possuem gênero gramatical de acordo com o componente que está mais associado à emergência do traço. Assim, algumas línguas baseiam-se mais em critérios fonológicos — como os segmentos terminais

dos itens —, enquanto outras baseiam-se mais em informações do componente semântico — como, por exemplo, os traços de sexo ou gênero social para referentes humanos em português e espanhol. Podem, ainda, se basear em outros traços morfológicos dentro do sistema, como as classes flexionais de línguas como o latim e o russo, em que nomes são classificados de acordo com as formas do paradigma para caso, gênero e número.

Casos em que elementos fonológicos dos itens correlacionam-se com as classes de gênero do sistema são considerados pelo autor como exemplares do que se classifica por "critérios fonológicos para atribuição de gênero". Corbett (1991) exemplifica esses casos mencionando o sistema altamente regular do hausa, língua afro-asiática falada na Nigéria. Nessa língua, palavras do gênero feminino infalivelmente terminam em *-aa*. Tal regularidade colabora para a atribuição do feminino a novos itens que adentram o sistema, quando estes terminam em segmentos percebidos como marca de gênero. Essa observação de Corbett sobre o hausa é relevante para nossas discussões sobre o papel da forma dos itens implementados no fenômeno da atribuição de gênero em português.

Os critérios semânticos para atribuição de gênero, segundo Corbett (1991), podem ser entendidos sob a generalização da associação conceitual. Esse mecanismo, presente em todas as línguas humanas e parte de nossa cognição, seria responsável por relacionar grupos de palavras sob uma mesma classe de gênero graças a suas propriedades semânticas, p. ex., o caso do português, mencionado acima, onde o gênero feminino é amplamente associado ao sexo feminino quando os referentes são animados. No entanto, o autor afirma que a atribuição de gênero a novos itens relacionada a critérios semânticos só se dá perante a ausência de fatores de natureza formal (fonológica ou morfológica), algo ainda disputado na literatura.

Neste trabalho, entendemos que, ao propor a generalização mencionada acima em relação a uma base semântica para o fenômeno da atribuição de gênero, Corbett deixa de considerar a complexidade dos diferentes critérios de base semântica potencialmente atuantes. Ainda, divergimos do posicionamento do autor quanto à sobreposição dos critérios de base formal sobre aqueles de natureza semântica. Nesse sentido, os resultados de nosso experimento com pseudopalavras corroboram com estudos que apontam maior influência dos fatores

semânticos, em relação aos formais, no fenômeno de atribuição de gênero a itens novos.

Além das chamadas regras normais (*normal rules*) para a atribuição de gênero, Corbett discute outros dois critérios já mencionados na literatura como importantes para o fenômeno. Segundo o autor, tais fatores devem ser considerados com cuidado, graças à sua natureza não sistemática: são critérios que se aplicam a casos isolados. É assim que o autor vê, por exemplo, o caso dos itens que adotam, por analogia semântica, o gênero de um item já presente na língua receptora. Portanto, quando Poplack, Pousada e Sankoff (1982) trazem à discussão o item *interview*, emprestado do inglês para o francês, e explicam seu uso no feminino como sendo orientado pela presença *a priori* de um item formal conceptualmente análogo no francês, para Corbett, não estariam em jogo regras sistemáticas.

O segundo caso não regular de Corbett (1991) é o que chama de “atribuição automática” (*automatic assignment*), que diz respeito às relações de marcação no interior de um sistema de gênero gramatical: quando o mecanismo de atribuição de gênero gramatical a um item novo não é capaz de computar regras formais que condicionem a escolha de um ou outro gênero, o gênero não marcado do sistema em questão seria atribuído ao item. Corbett discorda desse pressuposto, trazendo casos em que a explicação por critérios de natureza morfológica e fonológica substituem satisfatoriamente a atribuição por não marcação. Neste estudo, optamos por considerar a atribuição do não marcado como um fator independente, atuante na atribuição de gênero a itens novos em português. Contudo, entendemos o posicionamento de Corbett quanto a esse fator como compatível à visão aqui adotada. Ao assumirmos o feminino como gênero marcado da língua, são os casos em que ele é atribuído às novas palavras que devem ser explicados por critérios mais específicos.

Barkin (1980) apresenta questões que interessam à nossa discussão, principalmente pelo enfoque que se distancia, de certo modo, dos traços gramaticais internos aos itens, revelando outras características importantes do fenômeno. A autora leva em consideração características dos empréstimos e fatores periféricos do fenômeno de atribuição que, até então, haviam sido pouco considerados na literatura: a competência bilíngue dos falantes que utilizam os empréstimos, as fronteiras entre os dois sistemas por eles internalizados e os



diferentes graus de adaptação percorridos pelos empréstimos, desde o momento em que passam a fazer parte do novo léxico até o momento em que estão plenamente inseridos no sistema.

Barkin analisa os empréstimos do inglês em um *corpus* de língua falada composto de entrevistas com falantes de espanhol como primeira língua. Utilizando o sistema de classificação de Haugen (1953 apud BARKIN, 1980, p. 105), a autora classifica os itens levantados em (i) inteiramente adaptados, (ii) parcialmente adaptados e (iii) não assimilados.

A importância desses fatores para o fenômeno se mostra, por exemplo, quando a autora aponta para a tendência dos falantes de não usar determinantes no mesmo sintagma de itens ainda não assimilados ao sistema lexical da língua em uso, ou seja, substantivos que não passaram por processos fonológicos adaptativos e são irreconhecíveis por monolíngues como palavras da língua em questão, como em casos de *code switching*. Segundo Barkin, a tendência tem a ver com os níveis de consciência dos falantes em relação aos dois sistemas aos quais têm acesso. Evitar o uso de artigos e adjetivos com palavras totalmente alheias ao sistema, por exemplo, estaria relacionado à consciência dos falantes bilíngues quanto à ausência de gênero nominal no inglês.

Finalmente, a autora comenta a escassez de evidência para a atribuição de gênero operada por fatores formais — devida à baixa ocorrência de palavras terminadas em vogais átonas no inglês — e identifica as associações baseadas no componente semântico como aquelas que de modo mais regular fundamentariam a atribuição.

Kilarski e Krynicki (2005) analisam a atribuição de gênero a empréstimos do inglês no dinamarquês, no sueco e no norueguês. Os autores, a exemplo de Corbett (1991), dividem os critérios possivelmente atuantes na atribuição de gênero em dois grupos — formais e semânticos — executando uma análise minuciosa para determinar os critérios mais relevantes para os casos investigados. Essa análise dá-se através do modelo de função discriminante, que permite testar o grau de predizibilidade para atribuição de gênero em um dado item. Nesse modelo, são testados todos os critérios previamente alimentados e podem ser observados aqueles que se mostraram atuantes no emprego de um ou outro gênero.

O estudo busca conclusões em duas frentes: primeiramente, descobrir a importância relativa de diferentes critérios na atribuição de gênero dos itens vindos

do inglês nas três línguas; em segundo lugar, verificar até que ponto as especificações gramaticais desses itens determina, ou seja, é capaz de prever, seu gênero nas línguas receptoras. Para alcançar conclusões nessas duas direções, os autores estipularam um amplo conjunto de fatores variáveis que poderiam exercer um papel sobre a atribuição de gênero. Os 19 critérios escolhidos, em parte, referem-se a traços específicos das línguas germânicas do norte, por exemplo, seus sistemas de gênero bipartites, para o sueco e o dinamarquês, e tripartite, para o norueguês, sem correlações com a semântica dos referentes sexuados, como é o caso do português e do espanhol.

Além de critérios morfológicos dessa natureza, outros critérios morfológicos considerados foram o *status* derivacional dos itens, o sufixo e as ocorrências no plural e no singular. Já os critérios fonológicos incluíram polissilabidade, quantidade do último fonema, quantidade do penúltimo fonema, número de consoantes finais, presença ou não de parada glotal final (para o dinamarquês), entre outros. Finalmente, os critérios semânticos considerados identificavam os nomes como animados ou inanimados, concretos ou abstratos, pessoais ou não pessoais, e também de acordo com a presença de um equivalente na língua receptora, como, por exemplo, em português, o uso de *hot dog* em vista da palavra equivalente *cachorro-quente*.

Os resultados quantitativos revelaram que, em dinamarquês e sueco, o gênero comum (um dos dois gêneros desses sistemas) está super-representado no subconjunto dos empréstimos, ou seja, havia relativamente mais nomes de gênero comum dentre os empréstimos do que no repertório da língua como um todo. O modelo de análise discriminante aplicado corroborou o quadro, tendo selecionado o tipo de gênero como uma das variáveis relevantes para o fenômeno. Em relação à predizibilidade na atribuição de gênero gramatical a um dado item — um dos objetivos na utilização do modelo citado — uma regularidade meramente moderada foi verificada nos sistemas gramaticais analisados: o modelo tendo selecionado, no melhor cenário, 71.9% dos casos do dinamarquês.

O trabalho de Thornton (2009) sobre atribuição de gênero a empréstimos traz algumas das bases que utilizaremos para a análise dos empréstimos do inglês no português. Adotamos o modelo da autora para essa análise pelo seu poder descritivo. Os diferentes critérios semânticos por ela propostos dão conta da diversidade de casos observados no recorte de empréstimos do inglês.

A abordagem da autora se dá, inicialmente, sobre as bases da descrição de Corbett (1991), operando uma divisão entre critérios de natureza formal, de um lado, e semântica, de outro, para a atribuição de gênero a itens novos nas línguas. Analisando o comportamento de topônimos, neologismos e empréstimos do inglês no italiano, umas das principais contribuições da autora é o aprofundamento dos critérios de natureza semântica envolvidos na atribuição de gênero gramatical.

Thornton (2009) considera a analogia semântica como um conjunto de critérios que podem operar na atribuição ativa de gênero gramatical. Ao estudar certos grupos de palavras, a autora estipulou critérios semânticos menores potencialmente ativos no fenômeno: *gênero do hiperônimo, equivalente e associação*. No primeiro caso, quando um empréstimo entra em uma categoria conceitual maior já presente na língua receptora, é possível que adquira o gênero do substantivo genérico que rotula aquela categoria; assim, *uma yamaha* designa uma motocicleta (feminino), enquanto que *um yamaha* designa um instrumento musical (um teclado, masculino). Já a associação de conceitos equivalentes diferencia-se por, neste caso, a língua receptora contar com uma palavra que designa estritamente o mesmo elemento nomeado pelo empréstimo; enquanto que, naquele caso, frente à ausência do conceito no léxico mental dos falantes, um conceito associado é acionado e seu gênero projetado no empréstimo. Assim, a palavra *call*, usada no feminino em português, tem seu gênero atribuído através do acionamento de *chamada*, um equivalente tradutório; por outro lado, *ecobag*, um conceito novo, é usado no feminino diante do acionamento de *bolsa*, um conceito associado já existente no léxico mental dos falantes de português.

Segundo Haspelmath (2009), empréstimos introduzidos em uma língua a partir de outra, via de regra, possuem estruturas internas não analisáveis. Um caso que corroboraria esse pressuposto é descrito por Daniel (2018) em sua discussão do empréstimo do substantivo *mannerism* /'mænəɹɪzəm/, do inglês ao japonês, onde resulta na forma /ma.N.ne.ri.zi.mu/. O contraste entre a estrutura dessa nova formação e de sua base em inglês seria produto da adaptação fonológica à fonotática do japonês. Haspelmath menciona um processo fonológico posterior operado sobre o empréstimo, o truncamento, que teria reduzido a terminação /zimul — que na base *mannerism* corresponderia ao sufixo do inglês *-ism* — para /i/: /ma.N.ne.ri.zi/. Fundamentando o pressuposto sobre a estrutura dos empréstimos, o autor nota que, no japonês, processos fonológicos como o truncamento seriam

sensíveis à estrutura morfológica das palavras, o que impediria sua aplicação sobre sufixos, por exemplo.

A natureza estrutural dos empréstimos nominais é também discutida por Scher e Malagoli (2022), que, filiadas à DM, assumem a localidade de gênero na estrutura nominal ao modo de Kramer (2015). As autoras defendem que novas formações por empréstimo surgem na língua receptora como raízes não categorizadas, ainda que sempre derivadas de bases morfológicamente bem formadas da língua de origem. Como vimos no item 2.1.2.2.1 sobre a sintaxe de nominal e o traço de gênero, a abordagem de Kramer (2015), por exemplo, sob a MD, postula que um nome morfológicamente funcional deve conter uma raiz e um traço categorizador *n*.

Se, para Haspelmath (2009), empréstimos lexicais, enquanto novas formações, são estruturas desprovidas de estrutura morfológica interna, sob o escopo da MD sabemos que a derivação de um substantivo, por exemplo, exige a presença de um elemento categorizador ao lado de sua raiz. Assim, ao ser introduzido no PB, assume-se (como KRAMER, 2015, 2016; SCHER & MALAGOLI, 2022, entre outros) que um substantivo produto de empréstimo deve concatenar à raiz, derivada de um nome base na língua de origem, um elemento *n* provedor de propriedades funcionais essenciais: seu estatuto enquanto substantivo e especificações de gênero e classe temática. Como representado em (5), é somente após essa categorização no léxico do PB que poderá um empréstimo receber outros morfemas, como o sufixo de número.

Ainda, para empréstimos nominais com referentes humanos, o categorizador deve trazer uma especificação quanto à morfossintaxe de sua expressão de gênero, a qual seria, *i*[+FEM], que diz que o gênero atribuído é interpretável semanticamente.

Por mais que, como será apresentado na Seção 4, os empréstimos comuns de dois gêneros não apresentem, *a priori*, expressão fonológica de gênero, um substantivo como *influencer* dispõe de uma especificação em sua estrutura interna quanto à potencialidade morfossintática de expressar ambos masculino e feminino em PB. Essa peculiaridade é explicada por Kramer (2015, 2016) através dessa concepção de diferentes núcleos categorizadores *n*, que podem ser atribuídos a raízes a partir de propriedades do referente.

Esses aspectos do processamento dos empréstimos, que implicam a dependência da atribuição de um núcleo nominalizador *n* são inclusive apresentados como evidência para o tipo de categorização que, conforme defendem Scher e Malagoli (2022), se opera sobre essas raízes nos empréstimos nominais no coreano: “A principal evidência para defender que os empréstimos são categorizados como *n* [em coreano] vem da possibilidade de ocorrência de uma partícula de caso acusativo -eul/-reul junto a essas palavras, já que essas partículas são tipicamente anexadas a *n*, e não a qualquer outra categoria. (SCHER, MALAGOLI, 2022, p.122). No português, além do sufixo de número, há evidências para a classificação de uma nova formação enquanto substantivo pela posição sintática que passa a ocupar dentro das estruturas frasais. Em *O deadline foi estendido para o próximo mês*, subentende-se a classe de palavra de *deadline* por vir precedido por um artigo e precedendo um predicado, por exemplo.

Segundo Kramer (2015), como já apresentado, as diferentes configurações de gênero dos nomes do espanhol estão associadas a três diferentes núcleos nominalizadores *n*. Esse pressuposto importa à discussão dos empréstimos uma vez que implica a seleção de um traço específico durante a introdução desses substantivos ao léxico do PB.

Scher e Malagoli (2022) elencam alguns aspectos semânticos e fonológicos que são pertinentes ao se discutirem os empréstimos linguísticos. Em relação à semântica, as autoras afirmam que, embora se baseie em uma estrutura de raiz da língua de origem, o processo de formação de empréstimos faz com que surjam novas raízes, “inéditas”, na língua receptora, as quais podem diferir, e normalmente diferem, semanticamente das raízes-base do processo.

O aspecto fonológico é tratado de forma semelhante pelas autoras, sendo até certa medida independente da forma da palavra na língua de origem, uma vez que se realiza a partir das regras e padrões fonotáticos do sistema receptor, e que permite ainda que a nova forma passe por novos processos morfofonológicos, como a derivação do empréstimo *-delet-* em *deletar* e *deleção*. Por outro lado, nossa hipótese de que segmentos ou traços semânticos da forma de origem fundamentam em parte o processo de seleção de gênero, o que aqui seria visto como seleção de um *n* categorizador em particular, precisa relacionar os pontos em que se acessam certas propriedades semânticas do substantivo em sua língua de origem.

Burkholder (2018) aborda o que chama de “SDs mistos”<sup>25</sup>, produzidos por bilíngues de inglês e espanhol. Estruturas como *la house*, com um determinante do espanhol e um substantivo do inglês, por exemplo, são apresentadas como recorrentes na fala de bilíngues desse par de línguas, segundo a literatura sobre o tema, construções como *la house* chegam a ser dez vezes mais comuns que *the casa*, o que se explicaria pela (não) na assimetria de traços das estruturas nominais dos nominais de cada línguas. Enquanto que SNs do inglês manifesta nos diferentes nomes as propriedades de *número* — regularmente instanciado pela marca *-s* —, *pessoa* — nos pronomes *you, she, they; my, his, ours* etc. — e gênero, também e especial nos pronomes, como *her*[+FEM], *his*[-FEM], e de definitude, limitado aos artigos *the* [+definido] e *a/an* [-definido]; *expresso* pelos determinantes, a estrutura nominal do espanhol demonstra gênero, número e pessoa — este, nos pronomes pessoais. Por expressar uma propriedade funcional a menos, gênero, os determinantes do inglês são menos recorrentes em SNs na produção dos falantes estudados por Burkholder: enquanto *la house* o determinante expressa todas as informações do núcleo *house*, *the*, em *the casa* não comporta o traço de gênero inerente a casa.

Sendo os substantivos do inglês vazios para a categoria de gênero, se esperaria que, nos casos de SNs mistos de núcleo inglês e alvos de concordância do espanhol, a ausência de índices faria emergir o masculino em todos os casos. Nesse sentido, Burkholder (2018) faz menção ao fenômeno da atribuição de gênero por analogia com equivalente tradutório, similar ao que propõe o modelo de Thornton (2009) supracitado. Aqui entendemos essas caracterizações como atribuição de gênero por analogia semântica.

Segundo o autor, há três possibilidades para a atribuição de *ns* a substantivos do inglês em SNs em espanhol: i) pode se dar arbitrariamente; ii) pode haver uma preferência por *n*[-FEM] (assim como em Kramer 2015, 2016, o núcleo não especificado para gênero), levando à não marcação dos alvos e a emergência do masculino por *default*; iii) a atribuição se dá com base no nominalizador de um equivalente tradutório do espanhol, no caso em que estuda. Entendemos esse licenciamento das raízes de novas formações via seleção de um *n* de equivalentes tradutórios como uma instância do critério analógico para atribuição de gênero.

---

<sup>25</sup> *Mixed DPs.*

Burkholder traz à discussão ainda o cuidado teórico e metodológico a ser considerado por uma proposta de descrição do fenômeno. Nesse sentido, o autor questiona a adequação de se conceberem tais processos de analogia enquanto regras no fenômeno da atribuição de gênero a substantivos derivados interlinguisticamente. Embora reconheça haver um número considerável de estudos que apresentam evidências para seu papel, em uma ampla variedade de casos e muitas vezes com altas frequências de aplicação, o autor sugere que a analogia não deve ser considerada menos como regra e mais como uma tendência. Partindo de suas análises de corpora de produção em *língua-mista*<sup>26</sup>, apresentadas no mesmo estudo, esse tipo de analogia opera estritamente nos casos em que o falante consegue recuperar na memória um substantivo análogo e acessar corretamente sua informação de gênero, estando assim sujeita à imprevisibilidade.

Acrescentamos que a hipótese de atribuição de gênero por analogia com equivalentes masculinos não pode ser respaldada por análises de produção de modo geral, uma vez que sua atribuição pode sempre ser explicada pela *default*.

---

<sup>26</sup> *Mixed-language*.

### 3 EXPERIMENTO COM PSEUDOPALAVRAS

Assim como no estudo em Surreaux (2020), já descrito em 2.2.2, o instrumento descrito neste capítulo e a discussão dos seus resultados buscam elucidar o papel da forma de substantivos novos na escolha do gênero que lhes é atribuído por falantes de PB. A fim de aprofundar a hipótese anterior, focada no papel de terminações no processo, consideramos aqui também a atuação de segmentos iniciais que correspondem aos radicais de certos substantivos da língua.

Parte-se da hipótese de que a presença de segmentos recorrentes no léxico do PB, mais especificamente em substantivos femininos uniformes — aqui referidos como naturais —, pode estar associada a certa predizibilidade quanto a seu processamento para gênero. Segmentos terminais como *a*, *dade*, *ência* e iniciais como *batat*, *ciad* e *históri* poderiam assim atuar individualmente na marcação de gênero<sup>28</sup> de novas formações, o que seria esperado dado seu grau de associação com o feminino no léxico da língua — já que ocorrem em substantivos como *aluna*, *bondade*, *essência*; *batata*, *cidade* e *história*, respectivamente.

Justifica-se nossa escolha pelo método adotado — descrito em maior detalhe na sequência deste capítulo — por possibilitar à análise alguns ganhos quanto a conveniência e adequação. Dada a proposta de explicar o funcionamento do fenômeno da atribuição de gênero a novas formações e descrever as forças e unidades nele envolvidas, esse tipo de instrumento — bem como os descritos na seção 2.2.2 — permite certo controle sobre os contextos de aplicação do processo e o isolamento das unidades assumidas como relevantes. Aqui essas possibilidades se traduzem no uso de pseudopalavras que se apresentam aos participantes na forma de palavras que, assumimos, lhes sejam desconhecidas. Para isso, elaboraram-se as frases-veículo do instrumento — Já sabemos muito sobre \_\_\_ *falinara*, por exemplo, com estruturas que devem sugerir que as palavras testadas — nesse exemplo, *falinara* — sejam lidas como substantivos. Por sua vez, a instrução das questões, que pede que as lacunas sejam preenchidas com um artigo definido, *a* ou *o*, deve levar ao processamento morfosintático dessas palavras. Desse modo, nosso experimento simula a



introdução de uma nova formação nominal ao léxico do falante nativo de PB e o momento de seu processamento.

### 3.1 Metodologia

Elaboramos um experimento com pseudopalavras para testar algumas terminações recorrentes em substantivos femininos no PB: *ência*, *dade*, *a*, *ção* e *ão*, desta vez ao lado dos radicais *bol-*, *seman-*, *almofad-*, *biciclet-*, *cois-*, *históri-*, *cidad-*, *imagina[s]-*, *inspira[s]-*, *brincadeir-*, *laranj-*, *noit-*, *opini-*, *televis-*, *vacin-*. Cada um desses radicais e terminações naturais foram combinados a “radicais” e terminações artificiais<sup>2728</sup>, a fim de se isolarem os segmentos do português cuja associação com o feminino e a capacidade de suscitar sua atribuição tensionamos testar. Desse modo, elaboramos pseudopalavras como *laranjofe*, em que apenas a primeira parte da palavra é conhecida pelos falantes, e poderia fornecer informação de gênero, e *bolarição*, onde, por outro lado, a terminação é a única estrutura familiar.

A associação estatística dos sufixos testados com substantivos em femininos em português fundamenta-se em nossa pesquisa no Corpus do Português: NOW, que registra mais de 1 bilhão de palavras.

Tabela 4 – Ocorrência dos sufixos testados em palavras femininas

| Terminação | Número de <i>tokens</i> com as terminações no <i>corpus</i> | Substantivos femininos mais frequentes contendo os sufixos <sup>29</sup> |
|------------|---|--|
| a          | 178421679   | casa, polícia, forma   |
| ção        | 14096780  | relação, situação, ação  |
| ência      | 7263649   | agência, experiência, violência  |
| ão         | 3983090   | gestão, paixão, razão  |
| dade       | 3586563   | Cidade, responsabilidade,  |

<sup>27</sup> Para os fins da descrição do instrumento, entendemos segmentos e palavras naturais como aqueles atestados no uso da língua, os quais se opõem aos segmentos e palavras artificiais, empregados na elaboração das pseudopalavras do experimento.

<sup>28</sup> Por questões de praticidade, também nos referiremos aos constituintes das pseudopalavras como Raiz.art, Raiz.nat, Term.art e Term.nat; raiz artificial e natural, terminação artificial e natural, respectivamente.

<sup>29</sup> Ininteressantemente, o substantivo mais frequente terminado em *a* é o masculino *dia*, com 1.798.035 *tokens* registradas no CP:NOW.

Terminam em *ência* palavras como *demência*, *saliência*, *penitência*; em *dade* palavras como *integridade*, *objetividade*, *criatividade*; em *a* nomes como *barca*, *deusa*, *capoeira*; em *ção*, *ação*, *organização*, *maldição*; e em *ão*, substantivos femininos como *visão*, *lentidão*, *combustão*. As pseudopalavras e as estruturas consideradas em sua elaboração estão expostas na Tabela 4:

Tabela 5 – Estruturas para o teste de terminações

| <b>Terminações-teste</b> | <b>Pseudopalavras</b>                                       |
|--------------------------|---|
| <i>a</i>                 | <i>raca</i><br><i>taluz</i><br><i>falinara</i>              |
| <i>ência</i>             | <i>uzência</i><br><i>nafiquência</i><br><i>tolicanência</i> |
| <i>dade</i>              | <i>landade</i><br><i>matidade</i><br><i>garovidade</i>      |
| <i>ção</i>               | <i>vação</i><br><i>lareção</i><br><i>bolarição</i>          |
| <i>ção</i>               | <i>ruzão</i><br><i>calestão</i><br><i>valenião</i>          |

Os segmentos iniciais selecionados ocorrem em substantivos femininos frequentes no PB, de acordo com levantamento feito em nossos *corpora* de referência: estão entre os 10% substantivos mais frequentes do *Corpus do Português*, e entre os 500 itens dessa classe mais frequentes no *corpus* do Projeto VARSUL (Variação Linguística na Região Sul do Brasil), respectivamente. Crucialmente, esses radicais não formam nomes derivados ou compostos

masculinos que sejam recorrentes na língua (segundo nossos dois *corpora* de referência), como é o caso, por exemplo, do substantivo *casa*, que compartilha um radical com formas masculinas em uso: *casarão*, *caseiro*, *casebre* etc. A associação dessas raízes com formas masculinas poderia ser um fator perturbador em nosso experimento, uma vez que poderia suscitar a atribuição de masculino por um analogia com formas masculinas derivadas.

Além de considerarmos sua frequência, os 15 substantivos cujas raízes foram utilizadas no teste foram selecionados considerando-se os critérios formais de número de sílabas, acento e complexidade morfológica. Há cinco itens contendo três, quatro e cinco sílabas, seis são oxítonos e nove paroxítonos; cinco são derivados, e dez, simples. Dos derivados, dois são terminados em *ção*, dois em *ão* e um em *eira*. Dos simples, oito têm tema em *a*, e dois, em *e*.

Esses radicais foram combinados a terminações artificiais para formar as pseudopalavras a serem testadas no experimento. Elaboraram-se esses segmentos para que emulassem terminações nominais recorrentes em português. Assim como os sufixos *ção* e *ão*, seis das terminações são ditongos tônicos, três com cada um a das terminações *éi* e *êr*. Da mesma forma, seis dividem-se igualmente entre *ofe* e *ive*, tomando-se sufixos naturais também dissílabos e paroxítonos, como *dade* e *agem*.

Por fim, se elaboraram três pseudopalavras com radicais naturais seguidos da vogal átona final *e*. Esta terminação, embora bastante recorrente na língua, não se associa de forma relevante<sup>30</sup>, entendemos, à atribuição de feminino ou masculino, os tipos de gêneros considerados em nossa análise, o que possibilita, assim como nos casos dos pseudosufixos, isolar as porções relacionadas ao radical.

---

<sup>30</sup> A análise sobre o léxico do português, tanto em dados de uso escrito como falado, em Schwindt (2018) mostra que a átona final *e* se distribui de forma equilibrada entre o feminino e o masculino.

Tabela 5 – Estruturas para o teste de radicais

| Terminações | Substantivos base  | Radicais-teste     | Pseudopalavras       |
|-------------|--------------------|--------------------|----------------------|
| <i>e</i>    | <i>bola</i>        | <i>bol-</i>        | <i>bole</i>          |
|             | <i>semana</i>      | <i>seman-</i>      | <i>semane</i>        |
|             | <i>almofada</i>    | <i>salmofad-</i>   | <i>almofade</i>      |
| <i>ive</i>  | <i>bicicleta</i>   | <i>biciclet-</i>   | <i>bicicletive</i>   |
|             | <i>coisa</i>       | <i>cois-</i>       | <i>coisive</i>       |
|             | <i>história</i>    | <i>históri-</i>    | <i>historive</i>     |
| <i>êr</i>   | <i>cidade</i>      | <i>ciad-</i>       | <i>ciadêr</i>        |
|             | <i>imaginação</i>  | <i>imagina[s]-</i> | <i>imaginacêr</i>    |
|             | <i>inspiração</i>  | <i>inspira[s]-</i> | <i>inspiracêr</i>    |
| <i>ofe</i>  | <i>brincadeira</i> | <i>brincadeir-</i> | <i>brincadeirofe</i> |
|             | <i>laranja</i>     | <i>laranj-</i>     | <i>laranjofe</i>     |
|             | <i>noite</i>       | <i>noite-</i>      | <i>noitofe</i>       |
| <i>éi</i>   | <i>opinião</i>     | <i>opini-</i>      | <i>opiniéi</i>       |
|             | <i>televisão</i>   | <i>televis-</i>    | <i>televiséi</i>     |
|             | <i>vacina</i>      | <i>vacin-</i>      | <i>vacinéi</i>       |

Assim, pseudopalavras testadas, como *imaginacêr* ou *semanofe* devem remeter a substantivos do português derivados por sufixação. As terminações artificiais elaboradas compartilham com sufixos nominais naturais da língua pelo menos três características: padrão fonotático, tonicidade dos produtos e a forma como neles aparecem combinados aos radicais. Quando apresentados adjacentes a bases de substantivos que devem ser familiares aos participantes (como *bicicleta* e *bola*), assumimos que se reforça a analogia com sufixos do português (como numa palavra hipotética *\*bicicletagem*). Mas, crucialmente, ao contrário dos sufixos, os segmentos artificiais, naturalmente, não contêm conteúdo semântico.

O instrumento de teste compreende 60 palavras, naturais e artificiais, seguem quatro ordens de organização quanto a sua estrutura, que será: base natural com sufixo artificial, base artificial com sufixo natural, base e sufixo artificiais, e palavras naturais. Vinte substantivos compõem cada uma dessas quatro categorias, sendo que o grupo das palavras naturais, um terço do material, representa os itens distratores.

A variável tipo de sufixo, quando aplicada, tem como variantes cada uma das terminações testadas (*a*, *ção*, *ão*, *dade*, *ência*, *éi*, *êr*, *ive*, *ofe*, *e*). A variável acento foi classificada em três níveis, oxítonas, paroxítonas e proparoxítonas. A variável número de sílabas foi classificada em quatro níveis, e diz respeito ao tamanho do radical, que pode ser monossílabico, dissilábico, trissilábico e polissilábico, com quatro sílabas. Além disso, após a realização do teste, cada item foi classificado de acordo com o gênero que lhe for atribuído pelos participantes.

Em relação às variáveis sociais, foram coletadas, na primeira parte do questionário, dados acerca da idade, grau de escolaridade, curso de ensino superior, línguas faladas além do português e tempo de residência no exterior dos participantes (estas três, se aplicáveis).

Os substantivos que integram o experimento serão apresentados aos participantes no questionário através de 60 frases-veículo, geradas a partir das combinações das 60 palavras com um template fixo. Um exemplo de questão do experimento está representado abaixo em (2):

(2) *Já sabemos muito sobre \_\_\_\_ semanofe.*

a) *a*

b) *o*

O experimento foi realizado por meio da plataforma Google Forms e divulgado em redes sociais e grupos de e-mail. O formulário permaneceu aberto para respostas por 72h a partir das 17h do dia 23 de maio de 2023, e contou com a participação de 133 voluntários.

Em um primeiro momento, foram oferecidos os campos de informações sociais de cada informante e, em um segundo momento, as 60 questões, como em (2), que pediam que fossem respondidas com apenas uma de duas opções, os artigos masculino (o) e feminino (a) do português. As duas alternativas — (2)a–b —

foram ordenadas de forma aleatória para cada vez que o teste era iniciado. A cada questão, somente uma das alternativas poderia ser escolhida, remetendo ao masculino ou a feminino.

### 3.2 Resultados

O experimento foi respondido por 133 participantes. Quanto às variáveis sociais, a maioria dos participantes tem idades entre 26 e 40 anos (54,1%), tendo o restante entre 15 e 25 (24,8%) e 40 ou mais (21,1%). Em relação ao gênero social com que se identificam, a maioria dos participantes escolheu como resposta feminino (67,7%), seguido por masculino (30,1%) e gênero não binário (2,3%). Quanto aos cursos de graduação, completo ou em andamento, dos participantes, a informação relevante para nossa metodologia é em relação ao curso de Letras, que representa 16,5% das respostas. Por fim, 89,4% dos respondentes declara falar outras línguas, sendo que desses, 16,8% fala inglês e espanhol, 5% fala inglês e francês, 6,7%, apenas o inglês e 10,9% declara falar três línguas, sendo uma delas o inglês. O restante fala duas ou mais línguas, como alemão e espanhol ou espanhol, francês e LIBRAS.

Nota-se que as 15 palavras naturais foram todas atribuídas a seu gênero correto no mínimo 99% das vezes, o que demonstra certa confiabilidade quanto à atribuição de gênero pelos falantes durante o experimento. Apresentamos aqui a média da atribuição de feminino às terminações controladas no experimento. Essas médias são operadas sobre os conjuntos de três palavras contendo cada um desses sufixos. Comentários sobre o contraste entre as palavras dentro desses grupos, que dizem respeito ao acento da raiz e ao número de sílabas, serão feitos posteriormente nesta seção.

Conforme mostra a Tabela 1 abaixo, das pseudopalavras terminadas em *-ência*, a grande maioria teve o feminino atribuído pelos falantes, o mesmo foi verdade para as terminadas em *-dade*, para aquelas em *-ção*, 90%, e, em menor grau, para as palavras terminadas em *-a*. Por outro lado, pouco mais de 10% das pseudopalavras do experimento terminadas em *-ão* foram assumidas como femininas pelos participantes.

Tabela 1 – Atribuição de feminino com base nas terminações

| Terminação | Atribuição de feminino (%) |
|------------|----------------------------|
|------------|----------------------------|

|       |      |
|-------|------|
| ência | 96,7 |
| dade  | 94,2 |
| ção   | 91   |
| a     | 90   |
| ão    | 13,8 |

Fonte: Autor (2023).

Com relação às palavras com terminações artificiais (-*ofe*, -*ive*, -*éi*, -*êr*) e -*e*, apresentamos a Tabela 2 abaixo com a relação das médias da ocorrência de feminino para cada grupo, com e sem raízes naturais — p. ex., arupam-se *uzência*, *nafiquência* e *tolicanência*, formadas por um segmento inicial artificial e uma mesma terminação natural *ência*.

Tabela 2 – Atribuição de feminino com nas raízes

| Estruturas        | Atribuição de feminino (%) |
|-------------------|----------------------------|
| Raiz.artif + -e   | 31,1                       |
| Raiz.nat + -e     | 68,2                       |
| Raiz.artif + -ofe | 34,8                       |
| Raiz.nat + -ofe   | 59,7                       |
| Raiz.artif + -ive | 36,4                       |
| Raiz.nat + -ive   | 58,1                       |
| Raiz.artif + -éi  | 24,6                       |
| Raiz.nat + -éi    | 58,6                       |
| Raiz.artif + -êr  | 19,8                       |
| Raiz.nat + -êr    | 57,4                       |

Fonte: Autor (2023).

Com relação à variável número de sílabas, nos importa o resultado referente àquelas palavras onde a raiz foi testada como índice de atribuição ao feminino. Uma vez que o número de sílabas dessas palavras corresponde a uma maior expressão das raízes testadas, por exemplo, de *bole* (como raiz do feminino *bola*) em

comparação a *brincadeirofe* (de *brincadeira*), imaginamos que, para esse recorte, a variável poderia ter algum efeito, o que não parece ser o caso, segundo a Tabela 3 abaixo.

Tabela 3 – Atribuição de feminino e número de sílabas de R.nat+T.artif

| Número de sílabas | Atribuição de feminino (%) | Exemplos                       |
|-------------------|----------------------------|--------------------------------|
| 2-3               | 55,6                       | <i>bole, noitofe</i>           |
| 4                 | 63,2                       | <i>historive, inspiracêr</i>   |
| 5                 | 60,9                       | <i>bicicletive, imaginacêr</i> |

Fonte: Autor (2023).

Ainda, ao compararmos os dois tipos de palavras-teste, sejam eles, os que possuem um radical natural e uma terminação artificial (R.nat+T.art) e os que possuem um radical artificial e uma terminação natural (R.art+T.nat), tivemos os resultados expostos na Tabela 4 abaixo:

Tabela 4 – Atribuição de feminino e número de sílabas de R.nat+T.artif

|                               | Atribuição de feminino (%) | Exemplos                  |
|-------------------------------|----------------------------|---------------------------|
| Pseudopalavras<br>R.nat+T.art | 59,5                       | <i>almofade, semane</i>   |
| Pseudopalavras<br>R.art+T.nat | 77,5                       | <i>falinara, matidade</i> |

Fonte: Autor (2023)

Por fim, as variáveis sociais dos participantes, *idade, escolaridade, gênero social e envolvimento com curso de Letras* não demonstraram qualquer relação com a variável linguística *atribuição de feminino*.

### 3.3 Discussão

Assim como no experimento anterior, em Surreaux (2020), os sufixos que entendemos como associados ao feminino no léxico mostraram-se atuantes na atribuição de feminino: *ência, dade, ção* e *a*. Neste recorte das pseudopalavras em



que os sufixos foram testados, no entanto, contrasta a terminação *ão*, que pode estar associada ao feminino por seu caráter homófono, servindo tanto como um sufixo em abstratos femininos como *visão* e *cisão*, como em concretos aumentativos masculinos (*cabeção*, *mulherão*).

Uma atribuição de gênero com base em radicais iria de encontro a algumas propostas apresentadas aqui no item 2.1.2.2 quanto ao sistema nominal das línguas ibéricas e o *locus* de gênero na estrutura interna dos nomes (HARRIS, 1991; ALCÂNTARA 2010; KRAMER, 2015, 2016). Ainda que divergentes em muitos aspectos, os modelos de Harris, Alcântara e Kramer assumem para as informações de gênero uma localização externa à raiz: segundo Harris, os traços de gênero e classe são elementos categorizadores atribuídos a substantivos e adjetivos durante a derivação; Alcântara propõe que gênero se deriva a partir de morfemas categoriais de *Classe Formal*, os quais se combinam às raízes antes do processamento morfológico; ainda, Kramer assume uma estrutura nominal que combina radicais a núcleos nominalizadores *n* contendo as especificações de gênero.

Ao assumirmos que as terminações listadas na Tabela 5 são vazias para o traço de gênero, a atribuição generalizada do feminino a pseudopalavras como *noitofe*, *semane* e *bicicletofe* deve ser explicada por índices depreendidos dos segmentos correspondentes a seus radicais. Assim, entendendo os núcleos nominalizadores  $n_{u[-FEM]}$ ,  $n_{u[+FEM]}$  e  $n_{i[+FEM]}$ , podemos explicar a marcação do feminino em alguns itens como uma projeção do *n* análogo aos segmentos testados. Por exemplo, o substantivo *laranja*, que possui em seu nominalizador *n* uma especificação  $u[+FEM]$  quanto a gênero, é responsável pela atribuição de um mesmo *n* à raiz da nova formação *laranjofe*, quando se dá seu processamento pelos falantes no ambiente do experimento nos casos em que é considerado um nome feminino.

#### 4 ANÁLISE DE *CORPORA* DE EMPRÉSTIMOS

Aqui são apresentados e discutidos dados de três *corpora* de empréstimos nominais do léxico do PB, etimologicamente derivados do árabe, de línguas tupi e do inglês<sup>31</sup>. Esses subconjuntos lexicais foram introduzidos no português pelo contato linguístico e influência dessas três línguas. O fato de serem tipologicamente distintas, e de tal influência ter ocorrido em três momentos diacrônicos diferentes, justifica, em parte, sua escolha na investigação.

Para além dos fatores mencionados, que nos permitem generalizações quanto aos sistemas das línguas de origem, os contextos históricos e a variação diacrônica do português, as três línguas investigadas estão entre aquelas que maior efeito tiveram na ampliação de seu repertório nominal e, de modo geral, de seu léxico. Dentre as notações relacionadas às línguas de origem das entradas, a versão eletrônica do dicionário Novo Aurélio Século XXI, fonte utilizada, mostra como mais recorrentes o árabe, o francês, o inglês, o espanhol e o tupi. Assim, é preciso ainda que se justifique o recorte aqui apresentado, de três, desses cinco idiomas de origem. Ocorre que o espanhol e o francês, enquanto línguas românicas, naturalmente dispõem de estruturas, em geral, mais semelhantes às do português. Aqui, é crucial o fato de seus sistemas de gênero gramatical, bipartites (masculino e feminino), serem análogos ao do português. Uma vez que nossa intenção com a presente investigação, como será detalhado no parágrafo seguinte, é refletir sobre as regras formais e semânticas para a atribuição de gênero aos empréstimos, tal semelhança entre os sistemas de gênero impõe certa dificuldade de análise: a atribuição do mesmo gênero da palavra na língua de origem — como atribuir feminino em português a *omelete* por ser esse o gênero da base francesa — estaria competindo com demais regras, como aquelas de natureza formal e semântica.

Nosso objetivo com a análise proposta foi verificar possíveis evidências para processos de atribuição de gênero aos substantivos que teriam ocorrido no momento de sua introdução no português. Foram consideradas as regras de

---

<sup>31</sup> As palavras de origem árabe, em sua maioria, são assumidas como tendo origem da variedade da língua falada pelos ocupantes da Península Ibérica durante os séculos VIII e XIII, o árabe andaluz. A origem em línguas do tronco tupi, por sua vez, está relacionada a variedades mais hegemônicas da costa atlântica brasileira e dos pampas rio-platenses no decorrer dos séculos XVI a XVIII, como o tupinambá e variedades do guarani. Por fim, a maior parte dos empréstimos do inglês advém do inglês americano contemporâneo, havendo um pequeno conjunto de palavras adotadas da variedade britânica em meados do século XX.

atribuição de gênero da língua, descritas na seção 2.2, e a hipótese de sua atuação sobre propriedades semânticas e formais das bases dos empréstimos<sup>32</sup>.

Os dados relativos ao étimo árabe e tupi aqui analisados foram levantados do dicionário Novo Aurélio Século XXI, escolhido por oferecer ferramentas de busca por étimo e notações de gênero nas entradas. O *corpus* dos empréstimos do inglês, por sua vez, foi compilado através de levantamentos realizados por outros estudos sobre empréstimos do inglês em PB, o que se justifica por razão de, tanto nossa fonte mencionada acima quanto demais dicionários consagrados do português brasileiro não representarem satisfatoriamente essa influência no léxico dessa língua. Compreende-se que se deve esse fato ao caráter atual e à contínua introdução de empréstimos do inglês ao PB.

#### 4.1 Empréstimos do árabe

Os empréstimos do árabe ocorreram no português em sua maioria entre os séculos VIII e XV AD. Foram quase oito séculos de contato entre falantes de, entre outros, o árabe andaluz e falantes de variedades do latim vulgar então presentes na Península Ibérica que viriam a originar as línguas ibero-românicas modernas, como o catalão, o português e o espanhol. Aqui, em uma continuação do levantamento que apresentamos em Surreaux (2020), que teve por base 157 palavras, analisamos o gênero gramatical de 827 substantivos com essa origem.

De acordo com a classificação de Haugen (1956), a maior parte do léxico árabe implementado é representada por empréstimos culturais, denotando conceitos que não existiam na visão de mundo dos povos nativos da Península. São itens que representam novos bens de consumo, como laranja, azeitona e romã; tecnologias do domínio têxtil, como algodão, alfinete e almofada; e elementos comerciais, como armazém, açougue e aduana.

A nossa hipótese em relação ao recorte é que o critério de atribuição de gênero potencialmente atuante seria aquele relacionado à forma dos nomes do português, já que muitas das palavras em árabe terminam em segmentos que poderiam ser percebidos como *-a* átono final: *alfafa* e *alfazema* têm como formas de étimo *faṣfaṣa* e *al-ḵuzāmā*, respectivamente. Dado o distanciamento temporal em relação ao período da implementação da maioria dessas palavras, entendemos ser

<sup>32</sup> Aqui se faz uma diferenciação entre os empréstimos, os produtos do processo no léxico do português, e as suas bases, as formas ainda na língua de origem, que teriam sido processadas pelos falantes no fenômeno de atribuição de gênero durante sua adoção.

inapreensível o papel de critérios de natureza semântica para a atribuição de gênero a esses itens, como se observaram no recorte dos itens do inglês. É uma análise que exigiria a consideração do ambiente conceitual, ou seja, dos referentes que compõem o léxico dos falantes que participaram do contato linguístico na época em que ocorreu.

Há, no entanto, uma possível relação analógica entre os sistemas e as distribuições das marcas de gênero entre o árabe e o português. Nomes femininos, em muitas variedades do árabe, são fechados por um segmento vocálico átono -a, derivado da terminação -at, ainda atestável em algumas variedades e emergente enquanto resultado de fenômeno fonossintático no restante delas (CORRIENTE, 2008). Essa coincidência entre as gramáticas das línguas observadas traz alguns obstáculos para nossa análise. É possível, por exemplo, que o gênero dos empréstimos tenha sido deliberadamente projetado do gênero do árabe, por falantes da variedade românica e que, à época do contato linguístico em questão, possuísem algum conhecimento da língua emprestada.

Levantamos 827 substantivos do Novo Aurélio Século XXI, os quais continham a notação etimológica *ár*. Os itens levantados foram codificados de acordo com gênero (feminino, masculino ou comum de dois gêneros), segmento terminal (-a, -o, -e, ditongos ou consoantes) e grafema terminal do étimo, conforme notações da fonte consultada (o que inclui formas como <a>, <â>, <m>, <l> e clusters consonantais). A relevância desta última variável será abordada em detalhe abaixo.

Uma vez compilado e codificado o corpus, os itens reunidos foram analisados estatisticamente através do programa IBM SPSS Statistics, em vista da distribuição das variáveis e seus cruzamentos, em especial a comparação entre as ocorrências do gênero feminino, as terminações dos empréstimos em português e suas terminações nas formas de étimo. Ou seja, considerando-se o referido critério formal para atribuição de gênero, buscou-se identificar palavras que continham um segmento final análogo a -a em árabe e que tenham sido empregadas no feminino em português.

Conforme mostra a tabela a seguir, 60% dos empréstimos no recorte são masculinos, 37,8% são femininos e 13,3% são substantivos comuns de dois gêneros em português.

Tabela 5 – Gênero dos empréstimos do árabe

| Gênero                | %    | Exemplos               |
|-----------------------|------|------------------------|
| Masculino             | 57,4 | <i>Açougue, cifrão</i> |
| Feminino              | 36,2 | <i>Tarefa, safra</i>   |
| Comum de dois gêneros | 17   | <i>Refém, xiita</i>    |

Fonte: Autor (2023).

Quanto a seus segmentos terminais, temos 38,9% das palavras terminadas em *a*, 22,6% em *e* e 10,3% em *o*. Ainda, há um conjunto significativo de palavras que termina em consoantes como *l*, *r*, *s* (12%) e no ditongo *-ão* (5,6%), sendo que 10,6% dos substantivos termina em outros segmentos, como *ã*, *á* e *ei*.

Tabela 6 – Terminação dos empréstimos do árabe

| Segmento  | %    | Exemplos                 |
|-----------|------|--------------------------|
| <i>a</i>  | 38,9 | <i>alfafa, cifra</i>     |
| <i>e</i>  | 22,6 | <i>alambique, recife</i> |
| <i>o</i>  | 10,3 | <i>talco, xarope</i>     |
| <i>ão</i> | 5,6  | <i>alazão, sultão</i>    |
| consoante | 12   | <i>tambor, alcaçuz</i>   |
| outro     | 10,6 | <i>alquicé, xará</i>     |

Fonte: Autor (2023).

Dado nosso pressuposto quanto à marcação do gênero feminino e a regra de atribuição de gênero masculino por *default* possivelmente atuante na formação dos empréstimos nominais, nosso foco aqui, e nas análises a seguir, foram os casos das palavras classificadas como femininas. Desses 313 substantivos, conforme apresentado na tabela abaixo, uma parte expressiva termina em *a* átono, enquanto os outros se dividem entre *a* tônico e outros segmentos, como *e*, *s*, *l*, *ão*.

| Segmento | %    |
|----------|------|
| a        | 93,3 |
| á        | 3,4  |
| outros   | 3,3  |

Fonte: Autor (2023).

Nosso objetivo ao analisar conjuntos lexicais como este, como mencionado no início desta seção, é tentar explicar o fato de que alguns empréstimos foram adotados em português como substantivos femininos. Ao pressupormos que a ausência ou baixa saliência de índices para atribuição de gênero leva à escolha do masculino, buscamos nas bases dos empréstimos, ou seja, nas palavras ainda na língua de origem, informações que teriam fundamentado a atribuição do feminino em alguns casos. Aqui, entendemos que a estrutura dos étimos árabes que deram origem às formações em português, conforme proposto pela fonte lexicográfica utilizada, oferece possíveis evidências quanto aos contextos e à atuação de processos de atribuição.

Portanto, integram nosso *corpus* de substantivos emprestados do árabe, o étimo de cada um dos itens, como *samt* para *zênite* e *al-kaabaar* para *alcachofra*, tendo sido codificados seus segmentos terminais, pelo exemplo destes dois itens, -*mt* e -*aar*, respectivamente. Voltando-se apenas aos 292 empréstimos femininos em -*a*, temos que 84,3% deles possuem étimos fechados pelos grafemas ⟨a⟩ e ⟨â⟩, que, por representarem terminações vocálicas médio-abertas do árabe, podem ter sido associados à vogal átona final *a* do português no processo de atribuição de gênero. É o caso de palavras como *argola*, derivada de *al-gulla*, e *almofada*, derivada do árabe *al-mukhaddâ*. Os outros 15,7% dos empréstimos contêm étimos terminados em <m>, <q>, <bd> etc, como *tripa*, de *tarb*, e *zeduária*, de *zaduar*.

Conclui-se que a coocorrência aqui apontada de determinados grafemas dos étimos árabes e o segmento -*a* em empréstimos femininos no português estaria envolvida em fenômenos de atribuição de gênero ocorridos em momentos distantes da diacronia da língua. Além de haver evidências para a existência de um processo análogo sincronicamente ativo no português, referenciado nas discussões sobre aquisição e experimentos com falantes, é possível ainda propor que um fenômeno

assim tenha se dado com empréstimos formados a partir do contato com diferentes línguas em outras etapas da história do português.

#### 4.2 Empréstimos do tupi

Segundo a revisão de Moreira (2016) sobre as diferentes definições de “brasileirismo” na tradição e na prática lexicográfica, o aspecto comum a todas as concepções é o de que a noção refere termos utilizados especificamente na variedade brasileira da língua, estranhos ao português europeu e a demais variedades. Esse conjunto vocabular compreende, em sua grande maioria, vocábulos que têm sua origem em línguas africanas e ameríndias que, de uma forma ou outra, entraram em contato com o português a partir do século XVI.

O objeto desta análise em particular compreende o subconjunto dos *brasileirismos* a que a tradição lexicográfica atribui origem tupi, de modo geral, através da notação etimológica *Do tupi* nos vocábulos. O que aí se convencionou como *tupi*, por sua vez, precisa ser melhor explicado.

Na terminologia da linguística filogenética moderna, o termo tupi designa um dos mais diversos grupos linguísticos nativos do continente americano, com mais de 70 línguas descritas individualmente, agrupadas sob os troncos arikém, puruborá-ramarama, mondé, tupari, juruna, munduruku, mawé-aweú e tupi-guarani (DIETRICH, 2010). Ainda assim, entende-se que a implementação do léxico oriundo de línguas tupi no português resulta, em sua maior parte, do contato entre esta língua e certas línguas da família tupi-guarani com que os portugueses tiveram contato, o que se deu a partir de sua chegada ao continente.

Considerada a língua das origens do Brasil, o hoje extinto tupinambá, ou tupi antigo, foi a variedade dominante entre os habitantes da costa atlântica brasileira. Rodrigues (1958) define o tupi antigo como uma fase que compreende os séculos XVI e XVII, cujas fontes documentais, entre outras, são as gramáticas do padre José de Anchieta e Luís Figueira. Foi a língua de diferentes grupos étnicos que habitavam desde os atuais estados de Santa Catarina até a Bahia, estendendo-se, até o século XVII, ao Maranhão no interior do continente, adentrando a Amazônia. O tupinambá foi a língua por que se deram os primeiros contatos entre nativos e portugueses após sua chegada, tendo sido a língua das primeiras gerações de colonos e mestiços nascidos no novo território (RODRIGUES, 1958).

Rodrigues (1958) aponta algumas distinções necessárias quanto à fonte do léxico tupi-guarani no português. Por um lado, o tratamento dos empréstimos tupi-guarani no léxico português brasileiro deve considerar a influência mútua do tupinambá e de outras línguas relacionadas. O guarani antigo, de modo semelhante ao que ocorria com o tupinambá, se colocava como língua dominante no amplo território que hoje corresponde ao Paraná, Santa Catarina, Rio Grande do Sul, Paraguai e a região nordeste da Argentina. Sua influência persiste no léxico regional do Sul do Brasil: *guri*, *guaipeca*,

Entende-se que a maior variedade ameríndia falada no Brasil, do momento de sua colonização, era o tupinambá, variedade do tronco tupi-guarani, extinta desde meados do século XVIII, e sobre a qual se basearam a língua brasílica, a língua geral paulista e a língua geral amazônica. Além disso, o guarani, também do tronco tupi-guarani, falado majoritariamente no sul do Brasil, Paraguai e nordeste da Argentina, foi, desde os primeiros esforços descritivos das línguas ameríndias, bastante documentado, mesmo que em geral por fontes hispânicas. Segundo Rodrigues (1958), quanto à influência no léxico do português, foi o tupinambá e as línguas que dele provieram que recebem papel de destaque.

Conforme nosso embasamento teórico quanto ao processo de atribuição de gênero a novas formações, interessa o que o autor traz sobre a tipologia das línguas tupi, em especial quanto a seu inventário vocálico e a presença da vogal final -a nas formas do tupi que embasaram empréstimos em português. Segundo Dietrich (2010), embora alguns finais vocálicos, como -ã e -ê, tenham formas cognatas variáveis dentre as línguas tupis, dado o distanciamento diacrônico de seu étimo (podendo ser encontrados seus morfemas cognatos *-ang//em*), a vogal átona final -a aparece quase idêntica nas distintas línguas da família tupi, e também nas formas de étimo (*mani'oka* > *mandioca*<sup>33</sup>).

Finalmente, as notações do Novo Aurélio Século XXI, e dos dicionários portugueses em geral, de vocábulos de étimo tupi não fazem distinção das diferentes línguas dessa família. O que ocorre é que alguns vocábulos trazem regiões distintas do Brasil, onde são usadas as notas como RO, AM, RS (na entrada *puava*). Ainda assim, assumimos que as semelhanças entre essas línguas, principalmente no que diz respeito a sua fonotática, e, em específico à presença

---

<sup>33</sup> Conforme notação da versão eletrônica do Novo Aurélio Século XXI.



invariável da vogal média-alta final -a, como mencionado, seja suficiente para o embasamento inicial da análise.

Operou-se a pesquisa por todos os vocábulos da fonte de referência cujas entradas continham a notação *Do Tupi*, compondo nosso *corpus* de análise de 2.544 substantivos em uma planilha. Para vocábulos com duas entradas, com variação por gênero, criamos duas linhas no *corpus*. Por exemplo, *carapanã* pode referir-se tanto ao mosquito, quando é substantivo uniforme masculino, quanto aos indivíduos da etnia carapanã, quando é substantivo comum de dois gêneros. O mesmo para os casos do que pode ser considerado variação de gênero de um mesmo item, com mesmo significado, como em *apereá*, que traz a classificação *m.* e *f.* para uma mesma designação. Entendemos que é o mesmo fenômeno que ocorre em palavras fora desse recorte, dos quais alguns exemplos frequentes são *avestruz, alface, mousse*.

Entendemos que as palavras que constam no dicionário consultado já são as formas processadas pela gramática do português, com a ausência de grafemas que representariam encontros consonantais não licenciados na fonologia do português, como /mb/ inicial, que é adaptado como /m/ ou /b/, como apontam outras fontes etimológicas das formas das línguas tupi (BORGES, 2010): maracá > mbaracá. Ainda assim, a presença do segmento final que é grafado como <a> e processado fonologicamente como /a/ não sofreu alterações significativas, o que é crucial para nossa hipótese.

Os 2.544 itens foram codificadas de acordo o segmento terminal — <a, á, ã, e, é, ê, i, m, o, ó, ô, r, u> — e com a classificação de gênero das palavras na língua portuguesa — masculina, feminina, comum de dois gêneros. Importante notar que a terminação átona final, que interessa à nossa investigação, é predominante no recorte. Quanto ao campo semântico do conjunto, nota-se também que, quase integralmente, o *corpus* é composto por palavras que designam plantas e animais da fauna e flora nativa do continente, contendo 2.378 (96%) vocábulos as notações *Bot.* ou *Zoo*.

Em relação a sua classificação de gênero, dos substantivos analisados, 53,5% são masculinos, 40,2% são femininos e 6,3% são comuns de dois gêneros. Dentre as diferentes terminações observadas, o final átono -a aparece em 43,5% dos itens, seguidos pela vogal tônica -á, em 13,7% dos itens, a terminação tônica -i, em 13%, -u tônico, 12,4%, -é, 5%, -ã, 3%, m, 2,8%, ó, 1,6%, ê, 1,5%, -e átono,

1,1% e outras terminações, como os ditongos orais -ew e -ow, -o átono e tônico, l, r e s somam 2% nos itens da amostra.

Tabela 7 – Gênero dos empréstimos do tupi

| <b>Gênero</b>          | <b>%</b> |
|------------------------|----------|
| Masculinos             | 53,5     |
| Femininos              | 40,2     |
| Comuns de dois gêneros | 6,3      |

Fonte: Autor (2023).

Tabela 8 – Terminações dos empréstimos do tupi

| <b>Terminação</b> | <b>%</b> |
|-------------------|----------|
| -a átono          | 53,5     |
| -a tônico         | 40,2     |
| -i tônico         | 6,3      |
| -u tônico         | 12,4     |
| -e tônico         | 5        |
| Outras            | 4,6      |

Fonte: Autor (2023).

Ao cruzarmos as duas variáveis expostas acima, temos que os 1.362 substantivos de origem tupi aos quais o masculino foi atribuído terminam, em sua maioria, em vogais tônicas (69,1%), como -u (21,4%), -i (20,3%), -a (19,9%) e -é (7,5%), seguido por -m (4,5%), -a átono (3,7%) e demais terminações (22,7%). Dos 1.022 substantivos femininos do recorte, 83,7% terminam em -a átono, seguido pelas vogais tônicas -a (6,4%), -i (3,2%), -ã (2,4%), -u (2,2%) e demais segmentos (1,9%). Por fim, os 128 itens comuns de dois gêneros distribuem-se entre terminados em vogais tônicas (44,3%), -a (24,7%), -e (13%) e demais segmentos (18%).

Tabela 9 – Cruzamento entre gênero e terminação

|                       | <b>Masculino</b> | <b>Feminino</b> | <b>Comum de dois</b> |
|-----------------------|------------------|-----------------|----------------------|
| <b>-a</b>             | 3,7              | 83,7            | 24,7                 |
| <b>vogais tônicas</b> | 69,1             | 14,2            | 44,3                 |
| <b>outras</b>         | 27,2             | 2,1             | 31                   |

Fonte: Autor (2023).

Ao observarmos os itens com a terminação átona *-a*, percebemos que as palavras do tupi eram consistentemente associadas ao feminino quando traziam essa terminação, com 77,4% femininas, seguidas por 17,1% masculinas e 5,5% comuns de dois gêneros.

Em alguns casos, certas terminações, ou bases de composição, não influenciam no gênero de palavras. São casos como o feminino terminado em *-i* araraticupi (arara+tucupi), o mesmo pode ser observado em alguns compostos híbridos, não incluídos no corpus, com base nuclear à esquerda do português, e à direita do tupi, como merreca-apaí, que leva o gênero feminino da primeira base.

Por outro lado, em alguns casos, a não atribuição do feminino a palavras em *a* parece ser motivada pelo fato de o segmento contendo a terminação pertencer à base não nuclear do composto, como em amorepinima m. e amorepixuna m., cujos núcleos são amoré e o gênero que se atribui é masculino, por não marcação, embora o elemento não nuclear tenha um índice formal: *a* final.

É possível ver a atuação de mais de um fator de atribuição dando origem à variação de gênero em determinados itens, como no caso de tubi, uma espécie de abelha, que é registrado com entradas no feminino e no masculino. Para o primeiro caso, a analogia semântica com o termo feminino abelha parece ter sido responsável pela atribuição do gênero marcado, enquanto que, no segundo, a ausência de índices formais, ao lado do critério de não marcação, deram origem ao uso masculino. Nota-se ainda que uma parte significativa de palavras femininas que não terminam em *-a* como sendo designativos para árvores, trazendo a palavra árvore em suas definições: aítá, araribá, cuma-uaçú, curupitã, guarantã, entre outras. Ainda assim, há casos em que índices de ambos os critérios de atribuição de feminino estiveram presentes no item e, ainda assim, o masculino foi atribuído, como em arajaíba, que designa uma espécie de árvore.

A atribuição por analogia com um conceito associado no português também pode estar relacionada a casos em que existiu um composto em um estágio inicial, como marreca-ananaí, para ananaí ou abelha-aramá, em abelha, casos registrados no dicionário. Embora não ocorrendo nas entradas do dicionário consultados, casos assim poderiam explicar o feminino em outras palavras que não terminam em -a, como embirataí e cabatã, cujas estruturas contêm os nomes femininos embira e caba.

Por outro lado, grande parte das palavras masculinas terminadas em -a designam peixes, contendo a terminação comum peva/peba o que pode competir com o critério de atribuição formal, que atribuiria o feminino (CORBETT, 1991)

Uma ilustração das potencialidades para cada forma e seus significados vem através de palavras com três acepções, uma com cada classificação de gênero, como capixaba, que pode ser masculino (“pequeno estabelecimento agrícola”), feminino (“planta da família das euforbiáceas”) ou comum de dois gêneros (relativo ao estado brasileiro do Espírito Santo).

A maioria dos substantivos comuns de dois gêneros, acompanhados no dicionário da notação S.2.g, são etnônimos, como arari, mundurucu, parintintim. O restante são substantivos deadjetivais, tapejara e boboca, terminados em -a. Para os adjetivos, embora excluídos do recorte deste trabalho, cabe registrar que manifestam essa potencialidade de algumas palavras terminadas em -a (p. ex., dentista, atleta) para serem utilizadas com substantivos comuns de dois gêneros. Adjetivos como caipora, piroca, membeca são semelhantes a substantivos presentes no corpus.

Para muitos empréstimos aos quais o feminino foi atribuído mesmo diante ausência da terminação associada, parece ter papel decisivo a analogia com conceitos associados do português no processo de atribuição de gênero, uma vez que muitos designativos para árvores, por exemplo, vêm a ser usados como substantivos femininos, ainda que fechados por -ã, -á, -í etc, em geral mais frequentemente masculinos.

Por fim, o recorte possibilita a observação dos casos em que o masculino era atribuído a substantivos terminados na vogal baixa -a, grande parte dos quais pôde ser relacionada à estrutura interna das palavras — compostas, cuja base não nuclear apresentava a terminação — ou a domínios semânticos específicos, como o dos termos que designam peixes.

### 4.3 Empréstimos do inglês

Os substantivos de origem inglesa em uso no PB são o foco da terceira e última análise a ser discutida nesta seção, que dá continuidade ao debate e, de algum modo, à análise apresentada em Surreaux (2020) e Surreaux e Schwindt (2022). Reitera-se aqui, de início, o já mencionado desafio metodológico imposto por esse objeto em particular sobre a discussão que se propõe. Embora seja adequado para o estudo de empréstimos de origens como aquelas discutidas nas seções 4.1 e 4.2 acima, o Novo Aurélio Século XXI nos ofereceria um recorte pouco representativo da influência corrente do inglês sobre o léxico do PB.

Acontece que os eventos de incorporação de palavras árabes e tupis no português tiveram início, meio e fim em momentos já transcorridos da história da língua, durante eventos específicos de contato linguístico. Por outro lado, a influência da língua inglesa sobre o português se dá sincronicamente e de forma contínua. A natureza recente desse fenômeno é ilustrada por Amorim, Baltazar e Soares (2017), ao referenciar o *Boletim da Sociedade da Língua Portuguesa*, que, em 1993, contava com 13 empréstimos do inglês na língua, ao passo que, no ano de publicação do seu estudo, listava mais de 800 desses itens.

Diante disso, tomamos como recorte para o presente estudo um *corpus* que se constitui tanto de substantivos de origem inglesa presentes em nossa fonte inicial, o Novo Aurélio Século XXI, como aqueles trazidos por outros trabalhos recentes que, ao se dedicarem ao tema dos empréstimos do inglês no PB, dispõem de *corpora* próprios. Tais estudos são Kennedy (1971), Cano e Prado (2006), Figueiredo (2010), Valadares (2013) e Orsi (2015), os quais apresentam, respectivamente, empréstimos de fontes diversas, da área da informática, do campo das gírias, de revistas de alta circulação no Brasil<sup>34</sup> e do âmbito da moda. Nota-se que todos os *corpora* utilizados contavam também com verbos, interjeições, adjetivos puros e advérbios, os quais foram excluídos, por razões de foco, do recorte deste trabalho.

Assim sendo, compõem nosso *corpus* 1.150 substantivos, os quais foram codificados, de acordo com o gênero que lhes foi atribuído em suas fontes, ao segmento terminal das palavras inglesas de base e o segmento terminal que

---

<sup>34</sup> Valadares realizou o levantamento dos empréstimos de origem inglesa das revistas Istoé, Veja e Época.

recebem em português. Além disso, em nosso pressuposto sobre o papel da marcação do feminino no processo de atribuição de gênero, identificamos, para os empréstimos que recebem esse gênero em português, equivalentes análogos da língua, os quais podem ter fundamentado regras de atribuição de caráter semântico.

Diferentemente do léxico árabe e tupi analisados, o inglês, em virtude de sua fonotática, não apresenta um repertório contendo vogais átona finais. Os únicos exemplos, dos mais de mil itens levantados, são *mídia*. Além disso, o levantamento de Kennedy, que considera também palavras já grafadas segundo as regras do português, relaciona-se com o fenômeno da epêntese vocálica do português quanto a palavras terminadas em segmentos não licenciados como /k/, /d/ e /p/. Assim, palavras como *beque* (de *back*), *lorde* (*lord*) e *clipe* (*clip*) são também consideradas. Mais do que isso, anotamos em nosso *corpus* as palavras com tais terminações como terminadas em /i/, o segmento epentético, em suas formas em português.

Em relação à distribuição dos gêneros dentro do recorte, constatamos que 84% das palavras recebem gênero masculino. Em relação a dados do inventário lexical do PB em dicionários, e mesmo em dados de uso, ambos descritos em Schwindt (2018), isso parece extrapolar um certo padrão de produtividade de gênero na língua: ao contrário do que constatou Schwindt (2018), em relação ao relativo equilíbrio entre palavras femininas tanto no léxico dicionarizado do PB quanto no uso no sul do Brasil, em nosso recorte observamos que o gênero se distribui de maneira relativamente equilibrada. Mais do que isso, os levantamentos de empréstimos do árabe e do tupi, também apresentam substantivos e masculinos em maior grau de igualdade: como vimos acima nesta seção, nomes femininos compõem 36,2% e 40,2% dos empréstimos para cada caso respectivamente.

Tabela 10 – Gênero dos empréstimos do inglês

| <b>Gênero</b>          | <b>%</b> |
|------------------------|----------|
| Masculinos             | 94,3     |
| Femininos              | 4,4      |
| Comuns de dois gêneros | 1,3      |

Fonte: Autor (2023).

Entende-se que a menor ocorrência do processamento de itens do inglês como femininos em português está relacionada à falta de contextos para as regras de atribuição, especificamente aqueles de natureza formal, em que bases estrangeiras terminadas com segmentos analisáveis a *-a* átono eram empregadas no feminino. Ainda assim, o feminino parece ter sido atribuído por bases formais em quatro casos neste *corpus*: *bermuda*, *mídia*, *sinuca*, *soap opera* e *interface*. Este último em uma analogia com a palavra feminina *face* em português, um caso de uma regra em que a analogia formal não se dá com a terminação *a*, mas sim com todo um item lexical da língua.

Tabela 11 – Segmento terminal dos empréstimos do inglês

| Segmento  | %    | Exemplos                                       |
|---|------|--|
| e epentético (finais não licenciados) <sup>35</sup> | 53,1 | <i>hip hop, internet, nerd, notebook</i>       |
| Consoante nasal                                     | 19   | <i>cameraman, meeting, nylon, paging,</i>      |
| /r/   | 15   | <i>mega-hair, player, performer speaker,</i>   |
| /a/ e /o/ átono                                     | 4,6  | <i>big data, bingo, portfolio, water polo,</i> |
| outros  | 8,3  | <i>paintball, pixel, revival, fake news</i>    |

Fonte: Autor (2023).

Deste modo, resta explicar o fenômeno da atribuição do feminino no recorte aqui discutido. Há, pelo menos, dois tipos de regra de natureza semântica envolvidos nesse processo: aquelas de natureza referencial, em que o feminino é atribuído com base no sexo ou gênero do referente, e aquelas de natureza analógica, em que o gênero de uma palavra associada, ou um equivalente direto, é projetado sobre o item que é incorporado como empréstimo. Abaixo listamos exemplos de casos que relacionamos a cada um desses processos.

<sup>35</sup> Conforme deduzimos que seja o caso para as formas de origem do inglês para a grande maioria dos falantes do PB.

i) Feminino atribuído por propriedades do referente: *barwoman, cover-girl, drag queen, hostess miss, lady, topmodel* etc.

ii) Feminino atribuído por analogia com um substantivo do português: *bike, clipart, ecobag, grapefruit, fakenews, hardnews, ecobag, touchscreen* etc.

Ainda, há aqueles casos que não se explicam adequadamente por nenhuma dessas regras, os quais entendemos que tenham sido adotados no feminino por analogias específicas operadas em seus contextos de adoção, mas que não são mais recuperáveis em uma análise como a nossa, sendo eles substantivos que entraram na língua há mais tempo, como os mencionados no trabalho de Kennedy (1971): *gangue, van, kombi*.

Destaca-se, por fim, que as palavras comuns de dois gêneros, como *hóquei, gamer* e *DJ* têm essa classificação de gênero herdada diretamente da semântica dos seus referentes<sup>36</sup>. Em sua totalidade, os empréstimos que assim classificamos dizem respeito a funções ou designações de seres humanos, que podem ser ocupadas tanto por indivíduos do gênero masculino quanto feminino e, portanto, por substantivos que expressem esses dois gêneros em sua morfossintaxe.

#### 4.4 Síntese do capítulo

Vimos nas três análises desta seção que os mecanismos de atribuição de gênero de português têm evidências em diferentes processos de empréstimos na língua, derivados do contato linguístico com línguas distintas e em diferentes fases da diacronia. Observa-se que a operação de regras formais ou semânticas está associada à disponibilidade de contextos para essas respectivas regras nos itens de base para os empréstimos. Ou seja, as variedades de árabe e do tupi que estiveram em contato com o português contêm em seu léxico pistas formais que podem ter sido utilizadas pelos falantes quando do processamento de empréstimos delas provenientes. Diferentemente, os substantivos do inglês, em consonância com o sugerido em Surreaux (2020) e Surreaux e Schwindt (2022), apresentam pistas escassas dessa natureza, restringindo-se o gênero feminino a um número reduzido de casos, possivelmente atribuído por motivações semânticas mais do que formais.

---

<sup>36</sup> Por se tratarem de traços interpretáveis, segundo a MD.





## 5 CONSIDERAÇÕES FINAIS

Esta dissertação se fixou no processo de atribuição de gênero gramatical a novas formações no português brasileiro, visando a elucidar o papel dos diferentes índices envolvidos no processamento e na produção dos substantivos da língua. Através de uma revisão abrangente da literatura sobre gênero gramatical, desde seu tratamento em gramáticas portuguesas mais antigas, passando pelas abordagens estruturalistas que primeiro descreveram as estruturas da língua portuguesa com o rigor da ciência moderna, até diferentes abordagens de viés gerativista, buscamos adequar nossos instrumentos de análise e alcançar uma compreensão mais profunda desse fenômeno complexo envolvendo a cognitivista.

Nossa investigação reforçou o que concluem outros trabalhos sobre o tema: a atribuição de gênero em português é um processo multifacetado que envolve fatores fonológicos, morfológicos e associações semânticas. Demonstramos que os falantes de português confiam em uma variedade de pistas para atribuir gênero a novas palavras, incluindo componentes de sua estrutura formal, características semânticas e a analogia com o repertório lexical.

Nosso estudo também destacou a importância de se considerarem os processos de empréstimo lexical na compreensão da atribuição de gênero em português. Demonstramos que os empréstimos lexicais do árabe, do tupi e do inglês podem fornecer informações valiosas sobre como os falantes adaptam palavras estrangeiras para se ajustarem ao sistema de sua língua nativa, incluindo como atribuem gênero a essas palavras.

Ainda, apresentamos um experimento psicolinguístico com pseudopalavras que testou índices de atribuição de gênero presentes na estrutura dos nominais: terminações comumente femininas como *dade* e *ção*, além de porções iniciais, correspondentes a radicais de substantivos femininos uniformes frequentes no léxico do PB: *seman-*, *históri-*, *biciclet-*.

Em suma, esta dissertação pode contribuir para a literatura sobre o tópico e para uma melhor compreensão de como o gênero é atribuído em português e de como pode variar em diferentes contextos. Nossas constatações têm implicações tanto para a linguística teórica quanto para aplicações práticas, como o ensino de línguas e o processamento natural de linguagem e, através dos *corpora* lexicais históricos aqui compilados, temos a expectativa também de colaborar com a

descrição histórica das línguas envolvidas, em especial do português brasileiro. Esperamos que nosso trabalho inspire novas pesquisas sobre esse complexo e fascinante tópico e leve a novas descobertas sobre a natureza da linguagem.

Por fim, novas propostas de estudo podem incluir aqueles que usam imagens ou índices visuais que são atribuídos aos gêneros; a etapa diacrônica do português em relação a esse sistema; e, principalmente, propostas bastante atuais como aquelas que exploram a emergência de um gênero adicional na língua.

## REFERÊNCIAS

- ALCÂNTARA, C. As classes formais do português brasileiro. **Letras de Hoje**, Porto Alegre, v. 45, n. 1, p. 5–15, mai. 2010.
- ALEXIADOU, A. Inflectional class and the nature of gender. **Lingua**, 114(7), 905-937. 2004.
- ALI, S. **Gramática da Língua Portuguesa**. Rio de Janeiro: Livraria Acadêmica, 1923.
- \_\_\_\_\_. **Gramática Secundária da Língua Portuguesa**. 14. ed. Rio de Janeiro: Livraria Acadêmica, 1931.
- ALKIRE, T.; ROSEN, C. **Romance languages: A historical introduction**. Cambridge, UK: Cambridge University Press. 2010.
- AMORIM, R., BALTAZAR, R., SOARES, I. The Presence and Influence of English in the Portuguese Financial Media. **International Journal of Society, Culture & Language**, 5(2), 49-59. 2017.
- 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.
- AUDRING, J. Gender as a complex feature. In: PACIARONI, T.; THORNTON, A. M.; LOPORCARO, M. (eds.). **Language Sciences**, v. 43, p. 5–17, 2014.
- BARBOSA, J. S. **Grammatica philosophica da lingua portugueza ou principios da grammatica geral applicados á nossa linguagem**. Lisboa: Typographia da Academia das Sciencias. 1822.
- BARKIN, F. The role of loanword assimilation in gender assignment. **Bilingual Review / La Revista Bilíngüe**, v. 7, n. 2, p. 105–112, 1980.
- BLOM, E., POLIŠENSKÁ, D., WEERMAN, F. (2008b). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. **Second Language Research**, 24, 297–331. 2008
- BORGES, A. G. Nomes indígenas no português do Brasil. **Boletim da Academia Galega de língua Portuguesa**, nº 3, 2010, pp. 59-76. Organizado por HERNÁNDEZ, Xosé Afonso.
- BURKHOLDER, M. Language mixing in the nominal phrase: Implications of a distributed morphology perspective. **Languages**, v. 3, n. 2, p 1-28, 2018.

CÂMARA Jr, M. **Princípios de linguística geral como introdução aos estudos superiores da língua portuguesa**. 4ª edição. Rio de Janeiro: Livraria Acadêmica, 1964

\_\_\_\_\_. **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.

CARVALHO, N. M. **Empréstimos lingüísticos**. São Paulo: Editora Ática, 1986.

CHOMSKY, N. **The minimalist program**. Cambridge, Mass: MIT Press, 1995.  
CLEGG 2010

CLEGG, J. H. Native Spanish speaker intuition in noun gender assignment. **Lang. Des.** v. 12, 5-17. 2010

COMRIE, B. Grammatical Gender Systems: A Linguist's Assessment. **Journal of Psycholinguistic Research**, v. 28, p. 457–466, 1999.

CORBETT, G. **Gender**. Cambridge: Cambridge University Press, 1991.

CORRIENTE, F. **Dictionary of Arabic and Allied Loanwords**. Boston: Brill, 2008.

CUNHA, C.; CINTRA, L. **Nova gramática do português contemporâneo**. 4. ed. Rio de Janeiro: Nova Fronteira, 2007.

DANIEL, D. **Clipping as Morphology: Evidence from Japanese**. University of Calgary, Calgary, AB, 2018.

DIAMOND, J. **Armas, germes e aço: os destinos das sociedades humanas**. Rio de Janeiro: Record, 1999.

DIETRICH, W. Nuevos aspectos de la clasificación del siriono y del yuki (tupí-guaraní). In: MESSINEO, Cristina, MALVESTITTI, Marisa, & BEIN, Roberto (eds). **Estudios en Lingüística y antropología: Homenaje Ana Gerzenstein**. Buenos Aires: UBA, Fac. de Filosofía y Letras, 2008, p. 39-48.

FEDDEN, Sebastian; CORBETT, Greville G. Canonical gender. **Journal of Linguistics**, v. 52, n. 3, p. 495-531, 2016. DOI: <https://doi.org/10.1017/S002222671500057X>.

HALLE, M; MARANTZ, A. Distributed Morphology and the Pieces of Inflection. In: HALE, K. L.; KEYSER, S. J. (eds.) **The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger**. Cambridge: MIT Press, 1993. p. 111–176.

\_\_\_\_\_. Some Key Features of Distributed Morphology. **Papers on Phonology and Morphology**, Cambridge, v. 21, 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. The analysis of linguistic borrowing. **Language**, v. 26, n. 2, p. 210–231, 1950.

\_\_\_\_\_ (Ed.). **Bilingualism in the Americas: a bibliography and research guide**. Tuscaloosa: American Dialect Society, 1956.

HYÖNÄ, J., & Lehtonen, P. (2010). The role of unconscious processes in language use. **Trends in Cognitive Sciences**, 14(9), 424-431. doi: 10.1016/j.tics.2010.06.010.

JACKENDOFF, R. **X-Bar Syntax: A Study of Phrase Structure**. Cambridge: MIT Press, 1977.

IORIO, D. The noun phrase in Kibembe (D54). **Newcastle working papers in Linguistics**. v. 17. 2011.

KENNEDY, J. The influence of English on the vocabulary of Brazilian Portuguese. *Hispania*, Vol. 54, No. 2, pp. 327-331, 1971.

KILARSKI, G.; KRYNICKI, G. Not arbitrary, not regular: the magic of gender assignment. In: DELBECQUE, N.; AUWERA, J.; GEERAERTS, D. (Eds.). **Perspectives on Variation**. Berlin: Mouton de Gruyter, pp. 235–250, 2005

KRAMER, R. The morphosyntax of gener. In: **Oxford studies in theoretical linguistics**. n. 58. Oxford. Oxford University Press. 2015.

\_\_\_\_\_. The location of gender features in the syntax. **Language and Linguistics Compass** 10(11). 661–677. 2016. DOI: <https://doi.org/10.1111/lnc3.12226>.

LEONI, J. P. R. *Grammatica Portugueza ou arte de bem escrever e falar a língua portugueza*. 2. ed. Lisboa: Imprensa Régia, 1838.

LEVY, Yonata. On the early learning of formal grammatical systems: evidence from studies of the acquisition of gender and countability. **Journal of Child Language**, v. 15, p. 179-187, 1988.

LIGHTBOWN, P. M., & SPADA, N. **How languages are learned** (5th ed.). Oxford University Press, 2020.

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-EMPRESIMOS-DA-LINGUA-INGLESA-NA.pdf. Acesso em: 17 jun. 2021

POPLACK, S. A variationist perspective on language contact. In: ADAMOU, E.; MATRAS, Y. (Eds.) **Routledge Handbook of Language Contact**. New York/London: Routledge, 2020. p. 46–62.

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

PRIBERAM. Dicionário Priberam da Língua Portuguesa, 2020. Disponível em: <https://dicionario.priberam.org/>. Acesso em: 8 out. 2020.

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

RODRIGUES, A. 1958b. Classification of Tupí-Guaraní. *International Journal American Linguistics* 24:1231-4. 1958

SCHER, A. P.; MALAGOLI, R. G. Empréstimos em Morfologia Distribuída: termos do inglês em coreano e português. **Cadernos de Squibs: Temas em estudos formais da linguagem**, v. 7, n. 2, p. 111-128, 2022.

SCHWINDT, L. C. Zeros na morfologia nominal portuguesa à luz da Optimal Interleaving Theory. **Revista virtual de estudos da linguagem – ReVEL** [Edição especial n. 5], 2011.

\_\_\_\_\_. Exponência de gênero e classe temática em português brasileiro. **Revista DELTA**, São Paulo, v. 34, n. 2, p. 745–768, 2018.

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

SURREAUX, P.; SCHWINDT, L. C. Marcação de gênero gramatical em formações novas em português brasileiro. **Estudos Linguísticos e Literários**, Salvador, n. 72, p. 390–414, 2022.

SOARES BARBOSA, Jerônimo. *Grammatica philosophica da lingua portugueza*. Lisboa: Oficina de Simão Thaddeo Ferreira, 1774.

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

VIARO, M. E.; TRINDADE, M. G. Relações entre terminação e gênero morfológico em Said Ali: o índice -l no português. **Línguas e Instrumentos Linguísticos**, Campinas, SP, v. 25, n. 49, p. 198–218, 2022. DOI: 10.20396/lil.v25i49.8668980.

TSIMPLI, I; HULK, A. Grammatical gender and the notion of default: Insights from language acquisition. **Lingua**, 137(1), 128-144. 2013



## APÊNDICE

### A) *Corpus de empréstimos do árabe*

|             |             |               |
|-------------|-------------|---------------|
| aba         | almorávida  | fulano        |
| abaci       | almotace    | fustete       |
| abácida     | almotolia   | gabão         |
| abádida     | almoxarife  | gabela        |
| abassi      | almuadem    | gafa          |
| abassino    | almucábala  | gafar         |
| abedale     | almude      | galanga       |
| abelmeluco  | alnaibe     | gandulo       |
| abelmosco   | alpercata   | garra         |
| abencerrage | alqueire    | garrafa       |
| abenuz      | alqueive    | gazel         |
| abroque     | alquequenje | gazela        |
| açafate     | alquermes   | gázua         |
| açafrão     | alquibla    | gelba         |
| açaimo      | alquicé     | gengibre      |
| açaimo      | alquiez     | gergelim      |
| acelga      | alquila     | ginete        |
| acém        | alquilé     | girafa        |
| acepipe     | alquimia    | guadameci     |
| acéquia     | alquitara   | guitarra      |
| acerola     | altitena    | hadii         |
| acéter      | alúa        | hadith        |
| achaque     | aludel      | haique        |
| acica       | alufá       | haragano      |
| acicate     | alvaiade    | harém         |
| acitara     | alvanega    | harmala       |
| açôfar      | alvanell    | haxixe        |
| acofeifa    | alvará      | hegira        |
| açoite      | alvaraz     | hena          |
| açorda      | alvarral    | hene          |
| acotéia     | alveitar    | homos         |
| açougue     | alverca     | huri          |
| açubá       | alvíssaras  | ifrite        |
| açúcar      | alvitana    | imame         |
| açucena     | alvoroço    | intifada      |
| açude       | alxaima     | isla          |
| adafina     | ama         | ismaelismo    |
| adail       | âmbar       | ismaelita     |
| adarga      | anaco       | ismailismo    |
| adarme      | anadel      | ismailita     |
| adarve      | anafa       | ium-el-ahad   |
| adelfa      | anafaia     | ium-el-arbaa  |
| adelo       | anáfega     | ium-el-djuma  |
| adialfa     | anafil      | ium-el-khamis |
| adibe       | andaime     | ium-el-sebt   |

|            |             |                |
|------------|-------------|----------------|
| adixá      | andrajo     | ium-el-tani    |
| adobe      | anexim      | ium-el-thaleth |
| adua       | anfião      | jaez           |
| aduada     | anil        | jarra          |
| aduar      | anoque      | jasmim         |
| adufa      | anta        | javali         |
| adufe      | anuduva     | jihad          |
| agomia     | arabi       | julepe         |
| aguazil    | araca       | julepo         |
| aiassari   | áraque      | jumada         |
| aiatolá    | aravia      | laca           |
| alá        | arbim       | lacrau         |
| alabão     | argamandel  | lápiz-lazúli   |
| alacir     | argel       | laranja        |
| alaçor     | argola      | latão          |
| alambel    | armazém     | leilão         |
| alambique  | aroeira     | lemanó         |
| alambor    | arrabalde   | lezíria        |
| alambre    | arrabil     | lilás          |
| alamia     | arraia      | lima           |
| alaqueca   | arrais      | limão          |
| alarde     | arrátel     | loque          |
| alarido    | arrebique   | macio          |
| alarife    | arrecabe    | madracal       |
| alardoça   | arrecife    | mafoma         |
| alarve     | arrequife   | magarete       |
| alaúde     | arriaz      | magazine       |
| alazão     | arricaveiro | magrebino      |
| albacar    | arribe      | mália          |
| albacara   | arropa      | mameluco       |
| albacora   | arrobe      | manchil        |
| albafar    | arroz       | mandil         |
| albarda    | arruda      | maometismo     |
| albarrã    | arsenal     | maquia         |
| albarrada  | arzenete    | marabu         |
| albarrão   | arzola      | marabuto       |
| albetoca   | assassino   | marafona       |
| alboque    | atabal      | maravedi       |
| albornoz   | atabaque    | marcassita     |
| alboroque  | atafal      | marfim         |
| albricoque | atafona     | marlota        |
| albudeca   | atalaia     | maroma         |
| albufeira  | atambor     | maromba        |
| alcaçaria  | atanor      | maronita       |
| alcácer    | atarracar   | marrão         |
| alcachofra | ataúde      | marroquino     |
| alcáçova   | atum        | marzipã        |
| alcaçuz    | auge        | máscara        |
| alcadale   | aval        | masmorra       |
| alcaguete  | avaria      | mastaba        |

|            |           |            |
|------------|-----------|------------|
| alcaide    | avelório  | mate       |
| alcalde    | axabeba   | matraca    |
| álcali     | aximez    | mazorro    |
| alcamonia  | axorca    | mesquinho  |
| alcanave   | azáfama   | mesquita   |
| alcânave   | azagaia   | metical    |
| alcândora  | azambujo  | mezereão   |
| alcânfora  | azar      | mimbar     |
| alcântara  | azebre    | minarete   |
| alcanzia   | azeche    | mirabe     |
| alcaparra  | azeite    | miramolim  |
| alcaravão  | azeitona  | mocafo     |
| alcaravia  | azêmola   | moçárabe   |
| alcaraviz  | azenha    | monção     |
| alcaria    | azerve    | morabita   |
| alcarraza  | azeviche  | morabitino |
| alcateia   | aziar     | mosleme    |
| alcatifa   | azimute   | moxama     |
| alcatira   | azinhaga  | muarra     |
| alcatra    | azinhavre | muçulmano  |
| alcatrão   | azougue   | muçulmi    |
| alcatrate  | azul      | mudéjar    |
| alcatraz   | azulejo   | muezim     |
| alcatruz   | azumbre   | mufti      |
| alcavala   | babismo   | múmia      |
| alcoceifa  | babucha   | musa       |
| alcofa     | badana    | muxarabie  |
| álcool     | bairro    | nababo     |
| alcorão    | baldar    | nácar      |
| alcorce    | balde     | nacibo     |
| alcorque   | baldo     | nadir      |
| alcouce    | baobá     | nafé       |
| alcova     | baque     | nafta      |
| alcoveto   | bar       | natrão     |
| alcrevite  | baraço    | názir      |
| alcunha    | barbaca   | nesga      |
| alcuza     | barda     | nochatro   |
| aldebara   | barregana | nora       |
| aldeia     | bateia    | nuca       |
| aldrabão   | bedem     | olibano    |
| aldrava    | beduíno   | olíbano    |
| alecrim    | beliz     | osga       |
| alefriz    | benjoim   | osmanli    |
| aletria    | berbere   | oxalá      |
| alfaalface | berinjela | papagaio   |
| alfadia    | bezoar    | paparraz   |
| alfafa     | bismela   | quermes    |
| alfageme   | bodoque   | quibe      |
| alfaia     | bolota    | quilate    |
| alfaiate   | bórax     | químico    |

|            |          |             |
|------------|----------|-------------|
| alfama     | borgo    | quintal     |
| alfambar   | burca    | rababe      |
| alfândega  | burnus   | rabadão     |
| alfanje    | caaba    | rabeca      |
| alfaque    | cabaia   | rabi        |
| alfaqueque | cabide   | rajabe      |
| alfaqui    | cabila   | ramada      |
| alfaraz    | cacife   | rás         |
| alfarda    | cacifo   | razia       |
| alfarém    | caciz    | realgar     |
| alfarge    | caco     | rebite      |
| alfarrábio | cádi     | recife      |
| alfarroba  | cadimo   | recova      |
| alfavaca   | cadoz    | refém       |
| alfazema   | café     | regueifa    |
| alfeça     | cafeta   | rês         |
| alfeire    | cáfila   | resma       |
| alféoa     | Cafiz    | retama      |
| alfena     | cafta    | rígel       |
| alfenim    | caimacão | rima        |
| alferes    | calândar | rincão      |
| alfim      | cáli     | roque       |
| alfinete   | calibre  | rosalgar    |
| alfitete   | califa   | rume        |
| alfitra    | canana   | safári      |
| alfobre    | cande    | safena      |
| alfola     | candil   | safrá       |
| alfolva    | candilar | ságena      |
| alfombra   | cânfora  | saguão      |
| alforje    | cansim   | salá        |
| alforma    | capuz    | salamaleque |
| alforra    | caravana | salepo      |
| alforreca  | Carmesim | saloio      |
| alforria   | carmim   | sambuco     |
| alfóstico  | cártamo  | sândalo     |
| alfoz      | casba    | sanefa      |
| alfrezes   | catre    | sarabatana  |
| alfridária | ceca     | saramago    |
| alfuria    | cecém    | sarraceno   |
| algália    | ceifa    | sena        |
| algame     | ceitil   | sheik       |
| algar      | celamim  | simum       |
| algara     | cendal   | sindi       |
| algaravia  | cenoura  | sirage      |
| algarismo  | cequim   | siroco      |
| algarroba  | cerome   | soda        |
| algarrobo  | Ceroula  | sofá        |
| algarvio   | cetim    | solimão     |
| algazarra  | chabã    | suaili      |
| álgebra    | chafariz | sucata      |

|              |            |           |
|--------------|------------|-----------|
| algema       | chifra     | sufi      |
| algeroz      | chirivia   | sultão    |
| algibebe     | chué       | sumagre   |
| algibeira    | chúmeas    | suná      |
| algodão      | cifa       | sura      |
| algol        | cifra      | surrão    |
| algoz        | civeta     | tabefe    |
| alguergue    | colcotar   | tabi      |
| alguidar     | coraixita  | tabica    |
| alheta       | cordovão   | tabique   |
| alicali      | cotão      | tabule    |
| alicate      | cubeba     | taça      |
| alicerce     | cuscuta    | taforeia  |
| alidade      | cuscuta    | tagarino  |
| alifafe      | dado       | tahine    |
| alimangariba | damasco    | taifa     |
| álime        | damasquino | taine     |
| aliofaina    | daroes     | taipa     |
| alizar       | delfin     | talco     |
| alizari      | dinar      | taleiga   |
| aljamia      | dirém      | talim     |
| aljaravia    | djallaba   | tàmara    |
| aljava       | djim       | tamarindo |
| aljazar      | drogomano  | tambor    |
| aljófár      | druso      | tara      |
| aljorce      | dulcadá    | tareco    |
| aljuba       | dulija     | tarefa    |
| aljube       | elemi      | tareia    |
| almácega     | elixir     | tarifa    |
| almádena     | embelecar  | tarima    |
| almadia      | emir       | tarrafa   |
| almadraba    | enxaca     | tauxia    |
| almadraque   | enxaqueca  | teracena  |
| almáfega     | enxara     | timbale   |
| almagre      | enxaravia  | toranja   |
| almainha     | enxarrafa  | tremoço   |
| almanaque    | enxávega   | tripa     |
| almandra     | enxeco     | tufão     |
| almanjarra   | enxerca    | turbito   |
| almanxar     | enxova     | turqui    |
| almarada     | enxoval    | uale      |
| almargem     | enxovia    | ulemá     |
| almarraxa    | escabeche  | úsnea     |
| almece       | esfiha     | uzífuro   |
| almécega     | espinafre  | vizir     |
| almedina     | estia      | xabã      |
| almeia       | estirra    | xácara    |
| almeirão     | estragão   | xadrez    |
| almeitiga    | falaca     | xairel    |
| almeizar     | falca      | xale      |

|            |            |            |
|------------|------------|------------|
| almenara   | falua      | xáquima    |
| almexia    | faluca     | xara       |
| almirante  | fanega     | xareta     |
| almíscar   | fanfarrão  | xariá      |
| almóada    | fanga      | xaroco     |
| almocadém  | faquir     | xarope     |
| almocafre  | faraz      | xaveco     |
| almocávar  | farda      | xeque      |
| almocela   | farroupo   | xeque-mate |
| almocreve  | fasquia    | xerife     |
| almoeda    | fateixa    | xiita      |
| almofaça   | fatia      | zaga       |
| almofada   | fatímida   | zagal      |
| almofala   | febra      | zamboa     |
| almofariz  | fedaim     | zarabatana |
| almofate   | felá       | zarcão     |
| almofia    | ferragoulo | zarco      |
| almofreixe | fez        | zedoária   |
| almofrez   | filaça     | zegri      |
| almogama   | filele     | zeneta     |
| almogávar  | folá       | zênite     |
| almojávena | forro      | zero       |
| almôndega  | fota       | zoina      |

## **B) Corpus de empréstimos do tupi**

|              |              |            |
|--------------|--------------|------------|
| aguarapondá  | pixuna       | ajuruetê   |
| aguaraquiá   | poaia        | apê        |
| andirá       | poca         | arapuê     |
| baguá        | pocema       | auaetê     |
| caá          | poqueca      | bererê     |
| curupiá      | poranduba    | caaetê     |
| guajá        | porangaba    | cacundê    |
| guajará      | pororoca     | caetê      |
| içá          | poruca       | cuietê     |
| mará         | potaba       | cumatê     |
| uaiá         | potiguara    | curuperê   |
| ubá          | potiúna      | curupetê   |
| muirapixuna. | prejereba    | curuquerê  |
| abacatuaia   | puava        | enapupê    |
| abaraíba     | puba         | guaraperê  |
| abaruna      | puba         | inhapupê   |
| abatinga     | puçanga      | ipê        |
| abiiba       | puçanguara   | irerê      |
| abiurana     | punga        | irerê      |
| abuna        | pupunha      | jabebiretê |
| abutua       | pupunharana  | jaguetê    |
| açacurana    | quataquiçaua | morerê     |
| açairana     | quetua       | mororê     |

|                  |              |                |
|------------------|--------------|----------------|
| acajucica        | quaxinguba   | mujanguê       |
| acamatanga       | quiçaba      | poraquê        |
| acangapara       | quicuca      | sariguê        |
| acapociba        | quijuba      | saruê          |
| acapurana        | quirana      | tapiretê       |
| acaraaia         | quirera      | tatuetê        |
| acaracuaíma      | ramarama     | tembê          |
| acaracuíma       | sabiacica    | tereterê       |
| acaraiacuaíma    | sabiapoca    | tietê          |
| acarapirambocaia | sabiaponga   | tuietê         |
| acarapitanga     | saboarana    | uiruetê        |
| acarapixuna      | saboeirana   | umbê           |
| acaraúna         | sacaca       | urupê          |
| acariçoba        | sacaibóia    | abacaxi        |
| acariúba         | sacanga      | abati          |
| acarijuba        | saçupemba    | acari          |
| acariquara       | saguira      | acunati        |
| acarirana        | saicanga     | acuti          |
| acarapitanga     | saíra        | agami          |
| açoiaba          | samambaia    | aguti          |
| acuaua           | sambaíba     | agutiguepe-obi |
| acuera           | sambiquira   | aibi           |
| acutimbóia       | sananduva    | aipi           |
| água             | sanhaçuíra   | airi           |
| aiapana          | sapipoca     | ajuri          |
| aiereba          | sapiranga    | amuri          |
| aijuba           | sapiroca     | anabi          |
| aipixuna         | sapopema     | anaiuri        |
| airoba           | saporema     | anani          |
| aiuara-aiuara    | sapucaia     | anari          |
| aiuruapara       | sapucairana  | aningapiri     |
| aiurucatinga     | sapupira     | apaiari        |
| aiurujuba        | sapuruna     | apiari         |
| ajuacora         | sapuva       | aracambi       |
| ajuba            | sarabiana    | aracapuri      |
| ajugaíba         | saracuíra    | araçari        |
| ajurana          | saracura     | aracati        |
| ajurucuruca      | saracutinga  | araguaguai     |
| ajurujubacanga   | sariema      | araguari       |
| amanaga          | saúba        | arapari        |
| ambaíba          | saúna        | arapeti        |
| ambira           | saúva        | araratucupi    |
| amburana         | sebereba     | arari          |
| ameiva           | sericóia     | arari          |
| ameraba          | seriema      | arauri         |
| amisaua          | serubuna     | aricuri        |
| amoipira         | serutinta    | ariri          |
| amorepinima      | sipaúba      | ariri          |
| amorepixuna      | siri-candeia | arupai         |
| andira           | siririca     | atarubaqui     |

|               |             |            |
|---------------|-------------|------------|
| andiroba      | siriruia    | ati        |
| anga          | siriúba     | atiati     |
| angapora      | soca        | aueti      |
| angelimpinima | soçoca      | avari      |
| anhanga       | soroca      | bacupari   |
| anhangapa     | sororoca    | bacuri     |
| anhanguera    | suaçuapara  | bacuri     |
| anhinga       | suaçupita   | baguari    |
| anhuma        | suaçutinga  | baiquiqui  |
| anhumapoca    | suçuaia     | bariri     |
| aniba         | suçuapita   | beri       |
| anicauera     | suçuarana   | biaribi    |
| anicavara     | sucupira    | biri       |
| aninga        | sucurijuba  | biribiri   |
| aningapara    | sucuuba     | birigui    |
| aniúba        | suindara    | boi        |
| anjuba        | sumaúma     | bojobi     |
| anu-coroça    | sununga     | huri       |
| anu-coróia    | suruba      | huriqui    |
| ápama         | sururuca    | buriti     |
| apapocuva     | sutinga     | buvuari    |
| apeíba        | suumba      | buxuari    |
| aperana       | taba        | caapoboxi  |
| aperema       | tabarana    | caami      |
| apiranga      | tabatinga   | caapi      |
| apoitaguara   | tabebuia    | caaueti    |
| apojitaguara  | taboca      | caaxi      |
| apucuitaua    | tacana      | cabixi     |
| apuirana      | tacibura    | cabixi     |
| arabóia       | tacipitanga | cacumbi    |
| araçaíba      | taçuíra     | cacuri     |
| araçanga      | tacuruba    | caiabi     |
| aracanguira   | taguara     | cajati     |
| araçarana     | taiaçuíra   | cajubi     |
| araçaripoca   | taioaba     | camamuri   |
| aracaroba     | taioca      | cambuci    |
| aracimbora    | taioca      | camuri     |
| araçóia       | tajabemba   | cangati    |
| araçuaiava    | tajacica    | canguaxi   |
| aramandaia    | tamacarica  | cangurupi  |
| arapabaca     | tamarana    | caparari   |
| arapoca       | tamarutaca  | capitari   |
| araponguira   | tamba       | capixingui |
| arapuca       | tambaíba    | cari       |
| araputanga    | tambetara   | casbopoxi  |
| arara         | tamboera    | casobi     |
| araracanga    | tamearama   | catauari   |
| ararama       | tametara    | cati       |
| ararambóia    | tanajura    | catumbi    |
| ararandéua    | tangapema   | caubi      |



|             |                |                  |
|-------------|----------------|------------------|
| araraúna    | tanibuca       | cauxi            |
| araraúva    | taniça         | cernambi         |
| arariba     | taoca          | conabi           |
| araroba     | tapejara       | corocoturi       |
| arataca     | tapema         | cujubi           |
| arataca     | tapera         | cujumari         |
| arataia     | tapiara        | cumati           |
| aratanha    | tapieira       | cunuri           |
| aratinga    | tapiíra        | cunuri           |
| aratubaia   | tapinambaba    | cupi             |
| aratupeba   | tapioca        | curi             |
| araúna      | tapiocaba      | curi             |
| aricuriroba | tapira         | curumi           |
| arigbóia    | tapirá-caiena  | embiri           |
| ariramba    | tapiranga      | guacapi          |
| arirana     | tapiraua       | guacari          |
| ariranha    | tapiriba       | guacuri          |
| arua        | tapiúba        | guainumbiapirati |
| ataúba      | tapiucaba      | guanandi         |
| ateréua     | tapuia         | guanumbi         |
| atinga      | tapuirana      | guarandi         |
| atiriba     | tapuísa        | guarapari        |
| auíba       | tapunhunacanga | guirarepoti      |
| aurana      | tapururuca     | guiraundi        |
| autuparana  | taquara        | gureri           |
| auva        | taquaruva      | guri             |
| babaquara   | taraguira      | guriri           |
| babunha     | tarapema       | ingaxixi         |
| bacaba      | tarapitinga    | ipecuati         |
| bacaca      | tarioba        | ipequi           |
| bacupua     | tarumã-tuíra   | itacurumbi       |
| baitaca     | tataíra        | iurumi           |
| baquara     | tatajiba       | jaborandi        |
| barajuba    | tatapora       | jabuti           |
| baraúna     | tatarana       | jabuti           |
| batinga     | tatarema       | jacamaici        |
| batuera     | tatera         | jacamarici       |
| batuíra     | tatuaíva       | jacami           |
| batuquira   | tatuapara      | jaceguai         |
| batuvira    | tatucua        | jaci             |
| baúna       | tatupeba       | jacuapeti        |
| beguaba     | tatuquira      | jacuruxi         |
| beijucuruba | tatuxima       | jaguarundi       |
| beijuteica  | tauoca         | janari           |
| beruanha    | taxira         | japacani         |
| biboca      | taxirana       | japani           |
| bicuíba     | temapara       | jaraqui          |
| bira        | tembequara     | jauari           |
| biriba      | tepacuema      | jauari           |
| bituva      | terereca       | jaupati          |

|                 |               |             |
|-----------------|---------------|-------------|
| bocaiúva        | tetequera     | jequi       |
| boçoroca        | teúba         | jequiriti   |
| bóia            | tiba          | jipi        |
| boicininga      | tibuna        | juati       |
| boicotiara      | tietinga      | juciri      |
| boipeva         | tigugera      | jucuri      |
| boiquatiara     | tijucupaua    | jupati      |
| boiquira        | tijupaba      | juquiri     |
| boiúna          | timbaúba      | jurupari    |
| borajuba        | timborana     | juruti      |
| botara          | timboúva      | lambari     |
| bracatinga      | tinguaciba    | mamuri      |
| brejereba       | tipacoema     | mandaguari  |
| breu-jauaricica | tipuca        | mandi       |
| brujarara       | tiquara       | mandubi     |
| bubuia          | tiquira       | mapati      |
| bubuituba       | tiriba        | marimari    |
| bujiguara       | tirrica       | matapi      |
| buraçanga       | titara        | matiri      |
| buraquara       | titinga       | matupiri    |
| burara          | toba          | maturi      |
| burarema        | tobajara      | mboi        |
| butiatuba       | tobatinga     | membi       |
| butuca          | tocandira     | mendubi     |
| caachica        | tonca         | miri        |
| caajacara       | topatinga     | mucuri      |
| caajuçara       | torocana      | muici       |
| caamanha        | tororoma      | muirapixi   |
| caamembeca      | tovaca        | muriqui     |
| caapomonga      | traçanga      | mutumpiri   |
| caaponga        | tracutinga    | mututi      |
| caapora         | traíra        | nambi       |
| caavurana       | trairambóia   | nhambi      |
| caaxarama       | trapoeraba    | nhamburi    |
| caaxira         | trapomonga    | nhandi      |
| caba            | triba         | oanani      |
| cabiúna         | truirapeva    | oiampi      |
| cabiúna         | trumbuca      | oiti        |
| caboroca        | tubiba        | ouricuri    |
| cabriúva        | tubuna        | pacoviacari |
| cabureíba       | tucanabóia    | paracari    |
| caçarema        | tucupipora    | paracaxi    |
| caçaroba        | tucura        | paracuri    |
| cacica          | tucuruva      | parari      |
| caçuiroba       | tuijuva       | parati      |
| caçunanga       | tuiroca       | pari        |
| caçunanga       | tuiupara      | pati        |
| caiarara        | tujuba        | paturi      |
| caíba           | tunga         | pauxi       |
| caiçaca         | tupinambarana | pequi       |

|              |              |           |
|--------------|--------------|-----------|
| caicanha     | tupixaba     | peri      |
| caiçara      | turivara     | pipi      |
| caiçuma      | turuna       | pipiri    |
| caicumana    | tuxaua       | pirauxi   |
| caipira      | uacanga      | piri      |
| caipira      | uacapurana   | piriquiti |
| caipora      | uacataca     | piriri    |
| caipora      | uaicima      | poti      |
| caipora      | uarurembóia  | poxeti    |
| cajuçara     | ubacaba      | preguari  |
| cajurana     | ubiraçoca    | puvi      |
| camaísma     | ubirajara    | quati     |
| camarambaia  | uca          | quiriri   |
| cambacica    | uca          | quiriri   |
| cambaxilra   | ucuquirana   | sabiapiri |
| cambeba      | ucuuba       | saci      |
| cambeba      | ucuubarana   | sagui     |
| cambéua      | uéua         | sambaqui  |
| cambeva      | uiara        | sarandi   |
| cambica      | uirapiana    | siri      |
| camboa       | uiráuina     | sucuri    |
| cambuba      | umarirana    | suiriri   |
| cambuquira   | umirirana    | surubi    |
| camburiapeva | ura          | taci      |
| camina       | urarirana    | tamari    |
| camiranga    | uraúina      | tamati    |
| camocica     | uricana      | tambafóli |
| campeva      | urindeúva    | tambaqui  |
| camurupeba   | urubamba     | tamburi   |
| cangapara    | uruçacanga   | tanguari  |
| cangoeira    | urucuíba     | tani      |
| cangurupeba  | urucuriiba   | tapicuri  |
| canhanha     | urucuuba     | tapiri    |
| canhanha     | urumbeba     | tapiti    |
| canhembora   | urundeúva    | taquari   |
| caninana     | urupema      | taquiri   |
| capara       | urupuca      | tauari    |
| capejuba     | urutaurana   | tauari    |
| capepena     | utuaba       | tauri     |
| caperiçoba   | utuapoca     | taxi      |
| capinima     | utuaúba      | taxuri    |
| capininga    | uvaia        | teitei    |
| capintinga   | vatinga      | tingui    |
| capivara     | vitinga      | tipi      |
| capixaba     | ximbaúva     | tipiri    |
| capixaba     | xipaia       | tipiti    |
| capixaba     | xira         | tipói     |
| capoeira     | xiririca     | tocari    |
| capoeira     | acará        | toropixi  |
| capororoca   | acaraparaguá | tubi      |

|               |             |              |
|---------------|-------------|--------------|
| capuaba       | acaraparaná | tubi         |
| caquera       | acaraparauá | tucupi       |
| caraíba       | acarapindá  | tucuri       |
| caraíba       | acauá       | tupari       |
| carainha      | aiaçá       | tupi         |
| caranambuuba  | aiará       | tupi         |
| caranha       | aiapuá      | tupi-guarani |
| carapanaíba   | aitá        | tupinaqui    |
| carapiaçaba   | ajajá       | tururi       |
| carapina      | ajará       | uacari       |
| carapitaia    | ajupá       | uaipi        |
| carapitanga   | amaná       | uamói        |
| caraúba       | amapá       | ubari        |
| carcanha      | ambuá       | ubi          |
| cariboca      | anacá       | uiramembi    |
| carimbamba    | anaiá       | uiramiri     |
| carindiba     | ananá       | uirari       |
| carioca       | anauerá     | umari        |
| caripetirica  | andá        | umiri        |
| caritana      | andirá      | uxi          |
| cariúua       | anhangá     | vacapari     |
| carnaúba      | anujá       | vacari       |
| caruana       | apapá       | xuri         |
| caruara       | apecuitá    | abaçai       |
| caruca        | apereá      | açai         |
| catambruera   | apereá      | acaraí       |
| catambuera    | apiacá      | agai         |
| catanduba     | aquarapondá | aí           |
| catanguera    | aquaraquiá  | ananaí       |
| catapora      | ará         | anuaí        |
| catimpuera    | araçá       | anuí         |
| catuaba       | araguaguá   | apai         |
| catuaba       | araguirá    | apicuí       |
| cauaba        | aramá       | apuí         |
| cauaiúua      | aramaçá     | apuruí       |
| caúna         | aramatá     | aquai        |
| cávia         | aramatiá    | araçai       |
| caxinduba     | arapapá     | aracuí       |
| cernambiguara | arapuçá     | araguaí      |
| cica          | arará       | auai         |
| cipó-suma     | ararapá     | auí          |
| cipoaba       | araribá     | auí          |
| cipotaia      | arauaná     | bracuí       |
| ciriringa     | araxá       | caacambuí    |
| coariúba      | ariá        | cajuí        |
| cocama        | ariauá      | cambuí       |
| coirana       | arimbá      | caracaraí    |
| coivara       | aruá        | carai        |
| congonha      | aruaná      | caranaí      |
| copaíba       | arumaçá     | carandaí     |

|              |            |              |
|--------------|------------|--------------|
| copiara      | arumará    | carataí      |
| corindiba    | atuá       | caravataí    |
| corindiúba   | aturá      | cauí         |
| coroanha     | aturiá     | cuí          |
| coroca       | avará      | cumaí        |
| coroca       | axuá       | cupuaí       |
| corruíra     | bacaiá     | curupuruí    |
| cotiara      | bacopá     | cururuí      |
| cotinga      | bacumixá   | cutipuruí    |
| crauçanga    | baguá      | embirataí    |
| criciúma     | batará     | guapuí       |
| crindiúva    | biguá      | guaraçaí     |
| crupiara     | bijupirá   | iaguarataí   |
| cucura       | biribá     | iataí        |
| cuera        | biurá      | ituí         |
| cuia         | boicaá     | ituituí      |
| cuiambuca    | boitataí   | jacareí      |
| cuiapeua     | borá       | jamacaí      |
| cuiapitinga  | bracaiá    | japicaí      |
| cuiarana     | butiá      | jataí        |
| cuíca        | caá        | jataí        |
| cuipuna      | caapená    | jubaí        |
| cuiúba       | caapiá     | juquirá      |
| cujuba       | caatiguá   | jutaí        |
| cumbaca      | caçuá      | marreca-apai |
| cupânia      | caicurá    | maruí        |
| cupira       | cajá       | mijuí        |
| cupiúba      | camaá      | muí          |
| cupuaçurana  | camaiurá   | nanauí       |
| curcurana    | camará     | paranaí      |
| curera       | camboatá   | picuí        |
| curiacica    | cambucá    | piracuí      |
| curica       | camuá      | pirá         |
| curicaca     | candiubá   | puruí        |
| curimbaba    | cangambá   | sacaí        |
| curuaia      | canganguá  | saí          |
| curuba       | cangatá    | socoí        |
| curuca       | canguá     | suruí        |
| curuca       | cará       | tapiaí       |
| curucucica   | caracará   | tapií        |
| curuera      | caraguatá  | tapiocuí     |
| curupira     | carajá     | tatuí        |
| cururubóia   | caraná     | tieteí       |
| cururuca     | carandá    | tucumaí      |
| cururucica   | carapiá    | tuí          |
| cutiaia      | carapirá   | uaçaí        |
| cutimandioca | carará     | macuripal    |
| cutipaca     | carará     | abanheém     |
| cutiúba      | cararapirá | acutipum     |
| embetara     | carauá     | amendoim     |

|              |             |               |
|--------------|-------------|---------------|
| embiara      | catacá      | apacanim      |
| embira       | catambá     | apecum        |
| embiratanha  | catingá     | apicum        |
| embiricica   | catuá       | araticum      |
| emiraúna     | cauaiá      | ariquém       |
| gaponga      | caxangá     | buranhém      |
| gapororoca   | coaxixá     | cabamoatim    |
| garera       | coroaracaá  | camotim       |
| gargaúba     | cruá        | camucim       |
| gapiúna      | crubixá     | camurupim     |
| graúna       | cuatá       | canjurupim    |
| grumixama    | cumandatiá  | capim         |
| grupiara     | cunupiá     | capixim       |
| guabiraba    | curatá      | catucaém      |
| guabiroba    | curauá      | cauim         |
| guaçatonga   | curimatá    | chupim        |
| guaçatunga   | curimbatá   | conambim      |
| guacucuiá    | curuá       | cويم          |
| guaíba       | curuatá     | cunhantaim    |
| guaipeva     | curubixá    | cutucurim     |
| guajabara    | curumatá    | embiraém      |
| guajajara    | cururuá     | emerenhom     |
| guajuba      | cutitiribá  | emiraém       |
| guajuvira    | cuxá        | guaranhém     |
| guandira     | embuá       | guaxinim      |
| guaparaíba   | araciuirá   | igaratim      |
| guapeba      | gambá       | ipecumirim    |
| guapeva      | grapirá     | iratim        |
| guapicobaíba | grauçá      | itapicuim     |
| guapira      | grauçá      | jamaxim       |
| guaporanga   | gravatá     | japacanim     |
| guaraçaíma   | grumatá     | japim         |
| guaracava    | guajá       | japucanimpium |
| guaracema    | guajará     | jatobá-mirim  |
| guaracimbora | guará       | jurumim       |
| guaraçuma    | guaraguá    | maracatim     |
| guarajuba    | guaraná     | mequém        |
| guarancinga  | guarapirá   | micuim        |
| guarapicica  | guaruçá     | mirim         |
| guararema    | guatapará   | mirim         |
| guarariba    | guirapereá  | moquém        |
| guaravira    | guiraquereá | mucuim        |
| guariba      | guriçá      | muçum         |
| guariba      | iacaíacá    | mutum         |
| guaricanga   | ibiranhirá  | nhenhém       |
| guaricema    | içá         | niquim        |
| guaricema    | igapará     | parintintim   |
| guariroba    | igupá       | piraém        |
| guariúba     | indaiá      | piraquém      |
| guaruba      | ingá        | pium          |

|                  |               |             |
|------------------|---------------|-------------|
| guaxima          | ingá          | piurim      |
| guaxindiba       | inhambucuí    | pixaim      |
| gueirana         | inhambuquiá   | potimirim   |
| guaxinguba       | ipecuacamirá  | puxicaraim  |
| guibuguira       | ipecupará     | quatimirim  |
| guimberana       | ipecutauá     | sambacuim   |
| guirá-acangatara | irá           | tapicuém    |
| guiraguaçuberaba | iratauí       | tapicuim    |
| guiratinga       | itá           | tem-tem     |
| guiratirica      | itapeuí       | tem-tem     |
| guitroba         | itapuá        | tenharim    |
| guricema         | iuçá          | terém-terém |
| gurijuba         | iurará        | torém       |
| gurixima         | jabá          | tucum       |
| iandibacaba      | jabá          | tujumirim   |
| iapiruara        | jabutá        | tupiniquim  |
| iara             | jacá          | uiramirim   |
| ibabiraba        | jacamincá     | urumutum    |
| ibaiariba        | jacapá        | aracução    |
| ibapocaba        | jacarandá     | bapo        |
| ibiboboca        | jacaratiá     | caboclo     |
| ibiboca          | jacundá       | carancho    |
| ibicara          | jacundá       | caro        |
| ibijara          | jaguareçá     | cundurango  |
| ibioca           | jandiá        | gaturamo    |
| ibira            | janitá        | gibatão     |
| ibirapiroca      | japá          | iebaro      |
| ibirapitanga     | japuçá        | jacatirão   |
| ibirarema        | japuruxitá    | jenipapo    |
| ibirataíba       | jará          | jenipapo    |
| ibiratinga       | jará          | macuco      |
| ibixuma          | jaracatiá     | manduricão  |
| ica              | jaraguá       | mutirão     |
| icacoré-catinga  | jareuí        | pacamão     |
| icanga           | jaticá        | pirão       |
| icica            | jatobá        | puço        |
| icicariba        | jejá          | tacacazeiro |
| igaçaba          | jequiá        | tamatião    |
| igara            | jequitá       | tambaco     |
| igaruana         | jequitibá     | tamoio      |
| igaruana         | jeribá        | tapuio      |
| imbaubapuruma    | jerivá        | tero-tero   |
| imbetiba         | juá           | tico-tico   |
| imbuia           | jucá          | tijuco      |
| inambuquiçaua    | jundiá        | tinhorão    |
| inhaca           | jupará        | tucano      |
| inhaíba          | jupíá         | tucano      |
| inhambuanhanga   | jupindá       | tuco-tuco   |
| inhambupixuna    | jurará        | uribaco     |
| intanha          | juruparipindá | xerimbabo   |

|              |             |               |
|--------------|-------------|---------------|
| ipecacuanha  | jururá      | acaratimbó    |
| ipecupinima  | macerá      | apotó         |
| ipeúna       | mairá       | arapopó       |
| ipuca        | manacá      | ariacó        |
| ipuera       | manapuçá    | ariocó        |
| ipupiara     | mangangá    | bendengó      |
| ipuruna      | mangará     | biroró        |
| irara        | maniçauá    | borocotó      |
| iricurana    | manicujá    | caaigapó      |
| iritinga     | mapará      | caaxió        |
| itaberaba    | mará        | caicó         |
| itacuruba    | marabá      | carapó        |
| itaoca       | maracá      | carijó        |
| itapeba      | maracá      | cipó          |
| itapecerica  | maracajá    | cipó          |
| itapiranga   | maracujá    | curimbó       |
| itapiúna     | marajá      | curió         |
| itaquatiara  | marapá      | embeaxió      |
| itaúba       | maricá      | gapuicipó     |
| itaúna       | marubá      | guiraró       |
| ituiopinima  | marupá      | icipó         |
| itupava      | matamatá    | igapó         |
| iuçara       | matupá      | inhambuxororó |
| ivirapema    | mbatará     | inimbó        |
| ivitinga     | membitarará | itororó       |
| jabarandaia  | meruquiá    | jacaió        |
| jabutia      | miuá        | miroró        |
| jabutiba     | mucajá      | mocó          |
| jabutibóia   | mucuracaá   | mocororó      |
| jabuticaba   | mutá        | mocotó        |
| jacamacira   | pacajá      | potó          |
| jacanarana   | pacará      | promombó      |
| jacareúba    | pacová      | sanharó       |
| jacarina     | pajurá      | sarapó        |
| jacina       | panapaná    | socó          |
| jacitara     | paraná      | socoró        |
| jacobina     | parapará    | tecó          |
| jacuacanga   | paraturá    | temiminó      |
| jacuanga     | paricá      | timbó         |
| jacucaca     | paritá      | uaioró        |
| jacucanga    | patiguá     | zoró          |
| jacumaíba    | paxicá      | angatecô      |
| jagoirana    | pequiá      | bariiçô       |
| jaguacacaca  | pereiorá    | caapor        |
| jaguacininga | periná      | canitar       |
| jaguamitinga | piá         | jacamar       |
| jaguapeba    | picuá       | jaguar        |
| jaguapitanga | pindá       | potiguar      |
| jaguapoca    | pirá        | tapir         |
| jaguara      | pirá-andirá | tucanivar     |



|                |              |                 |
|----------------|--------------|-----------------|
| jaguaraíva     | pirapucá     | urubu-caapor    |
| jaguarapinima  | piriguá      | ananás          |
| jaguatirica    | piririguá    | aaru            |
| jamburana      | piruá        | abiu            |
| jananaíra      | pitanguá     | abu             |
| janatuba       | pitauá       | açacu           |
| janaúba        | pixuá        | acaju           |
| jandaia        | poracá       | acanguçu        |
| jandaíra       | puçá         | acapu           |
| jandiparaíba   | quipá        | acaraçu         |
| janiparindiba  | quiriripitá  | acaramuçu       |
| januaíra       | rubixá       | acarapicu       |
| japana         | sabiá        | acarapucu       |
| japara         | sabiá        | acariaçu        |
| japecanga      | saburá       | acoaramuru      |
| japeraçaba     | saguá        | acu             |
| japiaçoca      | saguaritá    | acutipuru       |
| japuíra        | saitauá      | aguaçu          |
| jaquiranabóia  | samanguaiá   | aguaraibá-guaçu |
| jaracambeva    | samburá      | aiaraçu         |
| jaramataia     | saná         | ajuru           |
| jarana         | sanamaicá    | ajuru           |
| jarandeua      | sanhoá       | ajuruçu         |
| jaranganha     | sapinhanguá  | anapuru         |
| jararaca       | sapuá        | andá-açu        |
| jaribara       | saputá       | anu             |
| jarina         | saraquá      | anuu            |
| jaritacaca     | sarará       | apecu           |
| jaroba         | sarobá       | arabebéu        |
| jarsiúba       | sauá         | arabu           |
| jaruva         | saiuíá       | aracaju         |
| jataíba        | simanguaiá   | aracatu         |
| jatecuba       | simongoiá    | aracu           |
| jatuarana      | sirigoíá     | arapaçu         |
| jatuaúba       | suiruá       | araracangaçu    |
| jaturana       | surucuá      | arataiaçu       |
| jauaraicica    | surucuá-tatá | aratu           |
| jauarana       | sururucujá   | araxixu         |
| jauarataceua   | tabuiáiá     | arimaru         |
| jaupoca        | taiá         | aru             |
| jejuíra        | taiuíá       | atiuaçu         |
| jeneúna        | tajá         | babaçu          |
| jequirioba     | tajurá       | bacacu          |
| jequitirana    | tamanduá     | bacu            |
| jequitranabóia | tamatiá      | bacucu          |
| jereba         | tambá        | bacurau         |
| jeremataia     | tambatajá    | baiacu          |
| jetaicica      | tambetá      | baquerubu       |
| jetica         | tambuátá     | beiju           |
| jia            | tamburupará  | betu            |

|                |             |            |
|----------------|-------------|------------|
| jibóia         | tamuatá     | biru       |
| jipioca        | tananá      | boiçu      |
| jipouba        | tangará     | boiru      |
| jiquitaia      | tangará     | brecambucu |
| jirabana       | tangurupará | bucu       |
| jiribana       | taperá      | buçu       |
| jitirana       | tapiá       | buiuçu     |
| juapoca        | tapiçuá     | cabaçu     |
| juba           | tarubá      | cacajau    |
| juçana         | tatuá       | caicau     |
| juçara         | tembetá     | cairuçu    |
| juiponga       | teringoá    | caititu    |
| juma           | tracajá     | caju       |
| jumbeba        | tracuá      | camapu     |
| juparaba       | trapiá      | canapu     |
| jurema         | tuaiá       | candiru    |
| juricana       | tucujá      | canguçu    |
| jurimágua      | tungurupará | canguçu    |
| jurubeba       | tupiná      | capicuru   |
| jurueba        | tupinambá   | carajuru   |
| jurujuba       | uaçá        | caramburu  |
| jurumbeba      | uaiapuçá    | caramuru   |
| juruna         | uajará      | caranambu  |
| juruparipiruba | uapuçá      | caraparu   |
| jurupetinga    | uauá        | carapuçu   |
| jurupiranga    | ubá         | caraxixu   |
| jurupixuna     | ubiraquá    | cariacu    |
| jurupoca       | uçá         | caruaru    |
| juruva         | uirá        | cauçu      |
| juruviara      | uiratatá    | conairu    |
| juruvoca       | uiratauá    | cuandu     |
| jutuaúba       | uruá        | cuintau    |
| jutubarana     | urubucaá    | cumã-uaçu  |
| juuna          | uxicuruá    | cumbaru    |
| mabóia         | vacuá       | cunauaru   |
| macacaúba      | vorá        | cunduru    |
| macambira      | acapitã     | cupu       |
| maçaquara      | acarapuã    | curu-curu  |
| maçarana       | acauã       | curu-curu  |
| maçarandiba    | acauã       | curupu     |
| maçaranduba    | acumã       | cururu     |
| macaúba        | apacamã     | embiruçu   |
| macaxera       | arabutã     | enxu       |
| maipoca        | araquã      | enxu       |
| mairata        | arumã       | euauaçu    |
| mamangaba      | caapuã      | garajau    |
| mamorana       | cabapitã    | guabiju    |
| manaíba        | cabatã      | guabiru    |
| mandaçaia      | caboatã     | guacuru    |
| mandiguera     | camboatã    | guaiamu    |

|               |              |              |
|---------------|--------------|--------------|
| mandioca      | carapanã     | guajeru      |
| mandiúba      | carapanã     | guaju-guaju  |
| manema        | caratuã      | guapiruvu    |
| mangaba       | carimã       | guapurubu    |
| mangangaba    | carimã       | guarabu      |
| mangarataia   | cauanã       | guaracavuçu  |
| mangaua       | cauanã       | guaracu      |
| maniçoba      | cunhã        | guarapu      |
| manicuera     | cunhã        | guaribu      |
| maniuara      | curimã       | guaru        |
| maniva        | curupitã     | guarupu      |
| manjerioba    | gragoatã     | guatambu     |
| maparajuba    | guarantã     | guiramombucu |
| mapareíba     | guarantã     | guiruçu      |
| mapironga     | guaratã      | guriaçu      |
| maqueira      | guavatã      | guribu       |
| maracatiara   | guriatã      | içabitu      |
| marajoara     | guriatã      | igarçu       |
| maranduba     | iacaninã     | igaruçu      |
| marica        | iaupê-jaçanã | imbu         |
| maritataca    | inhambuxintã | imbucuru     |
| marupaúba     | ipuã         | inambu       |
| marupiara     | irapuã       | inambu       |
| matarana      | itã          | indaiaçu     |
| matataúba     | itacuã       | inhambu      |
| matintaperera | jaçanã       | ipadu        |
| matintaperera | jacarandatã  | ipecu        |
| maúba         | jacumã       | ipu          |
| mejuba        | juruparã     | iruçu        |
| membeca       | macaguã      | itacuru      |
| merendiba     | macauã       | itapicuru    |
| metara        | macumã       | itu          |
| mijuba        | maracanã     | jacaçu       |
| minduba       | maracanã     | jacapu       |
| mingaupitinga | matrinxã     | jacarearu    |
| mirindiba     | meuã         | jacu         |
| mixira        | muçuã        | jacuru       |
| mocitaíba     | mucunã       | jacurutu     |
| mojica        | muiraquitã   | jaguaramuru  |
| monguba       | nhanjaçanã   | jamacaru     |
| moponga       | pacuã        | jamaru       |
| moranga       | pamonã       | jambu        |
| moroba        | panapanã     | japicuru     |
| morubixaba    | papuã        | japu         |
| morupeteca    | paracanã     | jaraguamuru  |
| mucica        | periantã     | jebaru       |
| mucuíba       | perinã       | jebebraju    |
| mucuna        | piã          | jeiu         |
| muçununga     | picumã       | jequitiguaçu |
| mucuoca       | piratantã    | jerimu       |

|                |               |             |
|----------------|---------------|-------------|
| mucura         | pituã         | jeticuçu    |
| muçurana       | puã           | jirau       |
| mucuraxixica   | punã          | jucuru      |
| mucutaia       | purumã        | jucurutu    |
| muiracaua      | sarã          | jundu       |
| muiracutaca    | suruanã       | juru        |
| muiraira       | tapinhoã      | jururu      |
| muirajuba      | tarã          | macucau     |
| muirajuçara    | tarumã        | macucu      |
| muirapaxiúba   | taticumã      | macuru      |
| muirapinima    | tincoã        | mandacaru   |
| muirapiranga   | tucumã        | mandiguaru  |
| muirapuama     | tupã          | mandu       |
| muiraquatiara  | uacumã        | mandu       |
| muiraketeca    | ubatã         | marmaiacu   |
| muiratinga     | xauã          | mingau      |
| muirauíba      | urueu-uau-uau | muirapucu   |
| muiuíra        | acarapinxame  | mulungu     |
| muiúna         | agutiguepe    | mundurucu   |
| mumbaca        | aígue         | murucu      |
| mumbava        | aracane       | murucututu  |
| mumbuca        | arauíne       | murangu     |
| munguba        | catete        | muraru      |
| mungubarana    | cuperope      | mureru      |
| mupéua         | curare        | muru        |
| mupororoca     | enduape       | mutumiju    |
| muquirana      | guaxe         | nambiuvu    |
| muquirana      | ipoteuate     | nhambu      |
| murajuba       | itogapuque    | nhandu      |
| murianha       | jabutifede    | nheengatu   |
| muriçoca       | jirmate       | ocarucu     |
| muriquina      | macurape      | pacu        |
| muruxaua       | paranauate    | pacuçu      |
| murupita       | petume        | pacuguaçu   |
| muruxaba       | pirabebe      | paiauru     |
| mutuca         | pirapebebe    | panacu      |
| mutucacaba     | quepquiriate  | papeá-guaçu |
| mutumporanga   | saquirabiape  | parabiju    |
| mututuca       | tacape        | parambiju   |
| narandiba      | tacuatibe     | pararu      |
| nhambipororoca | tareroque     | parauaçu    |
| nhandiroba     | tucumanfede   | paru        |
| nhanica        | uirafede      | perau       |
| obacatuara     | vuarame       | periperiaçu |
| obarana        | abaité        | piau        |
| oca            | abaré         | picaçu      |
| ocara          | acaé          | piracambucu |
| oeirana        | aimoré        | piracururu  |
| oiticica       | aimoré        | pirambu     |
| omágua         | ajaré         | pirambucu   |

|               |              |             |
|---------------|--------------|-------------|
| opaba         | amanaié      | piraniampu  |
| ouirarema     | amoré        | pirapucu    |
| ourana        | anambé       | pirarucu    |
| paca          | andirapuampé | pirupiru    |
| pacapeua      | andrequicé   | pissandu    |
| paco-caatinga | antacuré     | pitu        |
| paco-seroca   | apairandé    | pixiricuçu  |
| paçoca        | apé          | poaçu       |
| pacova        | apué         | pratibu     |
| pacova        | aquapé       | puaçu       |
| pacuera       | aracambé     | puaçu       |
| pacupeva      | aramaré      | punaru      |
| pamonha       | araué        | purupuru    |
| pamonha       | araueté      | purupuru    |
| panema        | aré          | putumuju    |
| papariúba     | aricoboé     | quatipuru   |
| paraambóia    | arubé        | quiriru     |
| paracaúba     | aturiapompé  | sabacu      |
| paracutaca    | aboré        | sanhaçu     |
| paracuuba     | baquité      | sapo-cururu |
| paraíba       | barimbé      | sapuvuçu    |
| paraíba       | batité       | sarabatucu  |
| paramarioba   | bembé        | sararau     |
| paranabóia    | boré         | savacu      |
| paranambuca   | bororé       | sinimbu     |
| pararaca      | caaeé        | sorocabuçu  |
| parauara      | caburé       | suaçu       |
| pariparoba    | cacaué       | suaçupucu   |
| pataquera     | caeté        | surucucu    |
| patativa      | caiaué       | sururu      |
| pavuna        | caimbé       | tabaiacu    |
| paxíúba       | caimbé       | tabaréu     |
| peba          | canindé      | tacuru      |
| pecapara      | canindé      | taiaçü      |
| peconha       | caraipé      | tajabucu    |
| peitica       | caraxué      | tambetaru   |
| penanguba     | caribé       | tambicu     |
| pepéua        | caripé       | tanatau     |
| pepeva        | carumbé      | tanduju     |
| pepuíra       | catajé       | tapeacuaçu  |
| pereba        | catulé       | taperu      |
| perereca      | cauré        | tapicuru    |
| peripetunga   | cavacué      | tapirapecu  |
| peripitinga   | copé         | tapuru      |
| peripomonga   | cubé         | taquaruçu   |
| peroba        | cunpeté      | tararucu    |
| peteca        | curupé       | tatu        |
| petimbuaba    | grumaré      | tatuaçu     |
| petinga       | guaiambé     | tauaçu      |
| petitinga     | guainambé    | teiuçu      |

|                |              |             |
|----------------|--------------|-------------|
| peúva          | guaripé      | tembetaru   |
| piaba          | guaxupé      | téu         |
| piabanha       | guiramembé   | téu-téu     |
| piaçaba        | guiraxué     | tiju        |
| piaçoca        | igarapé      | timbu       |
| piapara        | igarité      | timbucu     |
| picapara       | imbé         | tipu        |
| picuaba        | indaié       | tnguaçu     |
| picuçaroba     | inhambuapé   | tovacuçu    |
| picuipeba      | inhambuapé   | tucanuçu    |
| pindá-siririca | itaimbé      | uaçacu      |
| pindacuema     | itararé      | uacapu      |
| pindaíba       | iuquicé      | uacariguaçu |
| pindapóia      | jabutipé     | uacu        |
| pindauaca      | jacapé       | uaru        |
| pindoba        | jacaré       | uatapu      |
| pindopeua      | jacatupé     | uauaçu      |
| pindorama      | jaguapé      | uaçu        |
| pinguaciba     | jaguaracambé | ubuçu       |
| pinima         | jaguaré      | uirapaçu    |
| pipira         | jereré       | uirapuru    |
| pipirioca      | juraté       | umbaru      |
| pipoca         | maipuré      | unau        |
| pipuíra        | mandubé      | uracaçu     |
| piquirá        | manimbé      | uraçu       |
| piquitinga     | manopé       | uru         |
| pira           | maraximbé    | uru         |
| piraaca        | maué         | urubu       |
| piraba         | mondé        | urucatu     |
| piraboca       | mondé        | urucu       |
| piracajara     | mugambé      | urucu       |
| piracanjuba    | mundé        | uruçu       |
| piracema       | muremuré     | uruçu       |
| piracicaba     | mururé       | ururau      |
| piracururuca   | pajé         | urutau      |
| piraguara      | pavacaré     | urutu       |
| piraiapeva     | pixé         | urutu       |
| piraíba        | poracé       | vapuaçu     |
| pirajaguara    | poracé       | viruçu      |
| pirajica       | quatiaipé    | xexéu       |
| pirajupeva     | quatimundé   | xiru        |
| pirambeba      | quicé        | xuru        |
| piramutaba     | quicé        | acaú        |
| piranga        | sacubaré     | agaú        |
| piranga        | sambaré      | angaú       |
| piranha        | sapé         | araiú       |
| piraoba        | saporé       | arataciú    |
| pirapanema     | sateré-maué  | atarau      |
| pirapema       | seiré        | atingaú     |
| pirapeuaua     | sumaré       | caapitiú    |

|             |            |           |
|-------------|------------|-----------|
| pirapitanga | tacaré     | cabaú     |
| pirapitinga | tacuniapé  | cabiú     |
| pirapuia    | tamaquaré  | çaçaú     |
| piraquara   | tapirapé   | cametaú   |
| piraquera   | tarapé     | caminaú   |
| pirarara    | tembé      | cariú     |
| piraroba    | tembé      | cubiú     |
| pirauaca    | tié        | cuiú-cuiú |
| pirera      | timburé    | cupuaçú   |
| piririca    | tocajé     | curimataú |
| piroca      | tremembé   | cuxiú     |
| pirucaia    | tremembé   | embiú     |
| pitaica     | tucunaré   | guaiú     |
| pitanga     | tupé       | ibijaú    |
| pitangatuba | uanambé    | jamaracaú |
| pitigaia    | uapé       | jaú       |
| pitimbóia   | uarubé     | mandupitú |
| pitinga     | uiqué      | maú       |
| pititinga   | uirapejé   | mundaú    |
| pitomba     | uiraxué    | pajeú     |
| pitombarana | xié        | picaú     |
| pituba      | xué        | piraú     |
| piúca       | zoé        | pitiú     |
| piúna       | abaetê     | tambiú    |
| pixirica    | aí-ibiretê | teiú      |

### C) *Corpus de empréstimos do inglês*

|                |              |               |
|----------------|--------------|---------------|
| abstract       | holding      | target        |
| ace            | hole-in-one  | tartan        |
| activewear     | hollerith    | tattoo        |
| agenda-setting | homepage     | taxi-girl     |
| agitrop        | home theater | teaser        |
| agribusiness   | horsepower   | tee           |
| agrobusiness   | host         | teenager      |
| aids           | hostess      | telemarketing |
| aileron        | hot money    | teleprompter  |
| airbag         | hot dog      | terabyte      |
| airbus         | hotspot      | terrier       |
| airglow        | house organ  | test drive    |
| alis           | humour       | thesaurus     |
| alien          | husky        | thriller      |
| allnews        | hyperlink    | ticket        |
| alltype        | iceberg      | tie-break     |
| amish          | impeachment  | tide-die      |
| ampersand      | inner stage  | timer         |
| antidoping     | inning       | time-sharing  |
| antidumping    | input        | timing        |
| antispam       | insight      | toffee        |

|                   |               |                  |
|-------------------|---------------|------------------|
| antispypware      | interview     | toner            |
| approach          | intranet      | top              |
| ascaplot          | irish coffee  | topless          |
| assembler         | iron          | top model        |
| assembly          | isopin        | top-spin         |
| assembly language | it            | touch screen     |
| baby              | jab           | trailer          |
| baby-beef         | jamboree      | trainee          |
| baby-doll         | jam session   | training         |
| baby-sitter       | jazz          | transponder      |
| backbone          | jazz-band     | traveler's check |
| background        | jeans         | traveling        |
| backhand          | jet lag       | t-shirt          |
| backing vocal     | jet set       | twin-set         |
| backlight         | jingle        | twist            |
| backside          | jitter        | u-matic          |
| backup            | job           | underwriting     |
| bacon             | jogging       | up               |
| badminton         | john bull     | upgrade          |
| BAL               | joint venture | upload           |
| banana boat       | joystick      | upper-cut        |
| banana-spit       | juke-box      | voucher          |
| bandleader        | jumping       | valet service    |
| barbie            | kart          | van              |
| barman            | karting       | varactor         |
| bartender         | kelper        | varicap          |
| barwoman          | kerning       | vending          |
| baseball          | ketchup       | videobook        |
| basic             | kickbox       | videogame        |
| batch             | kickerboxing  | videolaser       |
| beach soccer      | kilobit       | videomaker       |
| beagle            | kilt          | video-on-demand  |
| beat              | king          | videowall        |
| beatnik           | king-size     | vintage          |
| benchmarking      | kit           | voucher          |
| benday            | kitchenette   | wade-giles       |
| best-seller       | kiwi          | waffle           |
| betting           | klaxon        | waiver           |
| bias              | knob          | wakashan         |
| big               | knock-out     | walkie-talkie    |
| big-bang          | know-how      | walkover         |
| big-church        | krill         | warrant          |
| bike              | lady          | water closer     |
| biofeedback       | lan           | water polo       |
| birdie            | langley       | web              |
| bit               | lan house     | webcam           |
| bitter            | laptop        | webmail          |
| black-tie         | laser         | weekend          |
| blank verse       | layback       | welfate state    |



|                  |                |                |
|------------------|----------------|----------------|
| blast            | legging        | western        |
| blazer           | leasing        | white-collar   |
| blimp            | lead           | whig           |
| blizzard         | let            | wi-fi          |
| blockbuster      | lift           | winchester     |
| blog             | lifting        | workaholic     |
| bloody mary      | light          | workshop       |
| blow-up          | link           | wow            |
| blue chip        | lip            | writ           |
| blue jeans       | living         | wysiwyg        |
| blues            | lob            | yearling       |
| bluetooth        | lobby          | yuppie         |
| blue-ray disc    | lock           | zapping        |
| blue-ray player  | lockout        | zoom           |
| blush            | loft           | accountability |
| body             | log            | air bag        |
| body-board       | login          | apartheid      |
| body-boarding    | logoff         | babyliss       |
| boiler           | logon          | blog           |
| bold             | long-play      | boom           |
| book             | long-playing   | bullying       |
| book             | loop           | bunker         |
| bookmaker        | lounge         | cakepops       |
| bookmark         | macromarketing | casteller      |
| boom             | mailing list   | chef           |
| boot             | mainframe      | chip           |
| borderline       | make-up        | closet         |
| bottom           | making of      | coach          |
| bourbon          | manager        | coaching       |
| bowl             | maple          | mainstream     |
| boy              | marine         | marketing      |
| brainstorming    | marketing      | nerd           |
| brake light      | mark-up        | netbook        |
| brandy           | marshmallow    | off-label      |
| break-even point | maser          | peeler         |
| break-point      | mass media     | playground     |
| briefing         | master         | post           |
| broadcast        | match          | portfolio      |
| broadcasting     | match-point    | print          |
| broadside        | media          | pulp           |
| brousse          | media mix      | ranking        |
| brownie          | medicine ball  | rapper         |
| browser          | medley         | cosplay        |
| brunch           | meeting        | crash          |
| budget           | megabit        | drone          |
| buffer           | megabyte       | dumping        |
| bug              | mega-hair      | e-commerce     |
| buggy            | megapixel      | e-reader       |
| bullying         | megawatt       | expert         |
| bungee-jump      | menu           | funk           |

|                |                  |               |
|----------------|------------------|---------------|
| bungee-jumping | merchandising    | game          |
| bunker         | microship        | game          |
| busdoor        | milady           | rating        |
| bush           | milk-shake       | remake        |
| button         | minke            | resort        |
| bye-bye        | miss             | retweet       |
| by-pass        | mister           | show          |
| byte           | mock-up          | showbiz       |
| cache          | modem            | site          |
| caddie         | monhair          | slogan        |
| caddy          | motocross        | socialite     |
| call-back      | motor-home       | software      |
| call center    | mountain-bike    | soul          |
| call girl      | mouse            | squash        |
| camcorder      | muffin           | spread        |
| cameraman      | music-hall       | stand-up      |
| camping        | muskogee         | hall          |
| campus         | negro spiritual  | hippie        |
| canvas         | nerd             | insight       |
| canyon         | net              | junkie        |
| carbолоy       | network          | link          |
| cartoon        | new journalism   | status        |
| cash           | newsletter       | streaming     |
| cashmere       | newsmaking       | stress        |
| caster         | nobreak          | strike        |
| cashmere       | no man's land    | stripper      |
| catch          | nonsense         | talk-show     |
| catering       | notebook         | thriller      |
| caucus         | nurse            | timing        |
| center-forward | nursery          | top           |
| center-half    | nylon            | voucher       |
| challis        | off-broadway     | waterboarding |
| chambray       | office-boy       | wingsuit      |
| charleston     | off-off-broadway | alias         |
| charter        | offset           | array         |
| chat           | onset            | assembler     |
| check-in       | op art           | backbone      |
| check-list     | open market      | backup        |
| check-out      | outlet           | basic         |
| check-up       | output           | batch         |
| cheddar        | outsider         | bit           |
| cheeseburger   | over             | boot          |
| cherry         | overalls         | browser       |
| chintz         | overbooking      | buffer        |
| chip           | overdose         | bug           |
| chippendale    | overhead         | byte          |
| chips          | overnight        | clipboard     |
| chroma-key     | over-price       | cluster       |
| chutney        | over-time        | cobol         |
| clean          | oxford           | console       |

|                     |                 |                            |
|---------------------|-----------------|----------------------------|
| clip art            | paddock         | cursor                     |
| clipboard           | pageant         | desktop                    |
| clipping            | pager           | DOS                        |
| close               | paging          | elo                        |
| closed caption      | paintball       | enter                      |
| closet              | palmtop         | exabyte                    |
| close-up            | pancake         | fax-modem                  |
| clown               | paper           | gateway                    |
| club soda           | parking         | gigabyte                   |
| cluster             | pashmina        | hacker                     |
| cocker spainel      | patchwork       | hardware                   |
| cockpit             | pay-per-view    | help                       |
| coffee break        | pedigree        | hipermídia                 |
| commercial<br>paper | peeling         | hyperlink                  |
| commodity           | pellet          | input                      |
| common law          | pence           | interface                  |
| compound            | pen drive       | job                        |
| container           | perfect binding | joystick                   |
| cookie              | performer       | kilobit                    |
| copyright           | permafrost      | laptop                     |
| cottage             | permifrost      | login                      |
| countertrade        | pet shop        | loop                       |
| country             | pickup          | mainframe                  |
| courier             | pidgin          | master                     |
| cover-girl          | piercing        | modem                      |
| cowboy              | pinotage        | mouse                      |
| crack               | pint            | palmtop                    |
| cracker             | pin-up          | pixel                      |
| crawl               | pitboy          | plotter                    |
| cream cracker       | pit bull        | postscript                 |
| cree                | pitch           | prompt                     |
| creek               | pixel           | scanner                    |
| crooner             | plantation      | script                     |
| cross-country       | play            | setup                      |
| crossing over       | playback        | site                       |
| crowding-out        | playboy         | time-sharing               |
| crumble             | player          | upgrade                    |
| cult                | playground      | web                        |
| cup                 | plotter         | fashion<br>revolution day  |
| cutback             | plush           | fashion<br>revolution week |
| dancing             | pocket book     | fash mob                   |
| deadline            | podcast         | green carpet<br>challenge  |
| decanter            | point-break     | beleaf                     |
| default             | pointer         | fashio pact                |
| delicatessen        | poodle          | global fashion<br>agenda   |

|                |               |                          |
|----------------|---------------|--------------------------|
| delivery order | pool          | nylon                    |
| demarketing    | pop           | re-sale                  |
| derby          | pop art       | upcycling                |
| design         | pop-up        | fashion revolution       |
| designer       | postscript    | day                      |
| desktop        | pot-pourri    | corporate responsibility |
| despatch money | press-kit     | haulternative            |
| diesel         | press-release | startup                  |
| disco-music    | primer        | basquete                 |
| dimmer         | prime rate    | beisebol                 |
| display        | promoter      | beque                    |
| divot          | prompt        | bermuda                  |
| dolby          | prompter      | bife                     |
| dolly          | prospect      | bingo                    |
| donut          | pub           | boicote                  |
| doping         | publisher     | bote                     |
| download       | punch         | boxe                     |
| downsizing     | punching ball | breque                   |
| double-bogey   | punk          | bridge                   |
| double-face    | punk rock     | catch                    |
| double-faced   | putt          | chance                   |
| drag queen     | putter        | clip                     |
| drawback       | puzzle        | clube                    |
| drive          | q-bit         | copyright                |
| drive-in       | quark         | cocktail                 |
| driver         | quart         | corner                   |
| driver-thru    | rack          | cowboy                   |
| dry-framing    | rafting       | dancing                  |
| dumping        | ragtime       | destrôier                |
| duocake        | railway       | detetive                 |
| duty-free shop | ranking       | disc-jockey              |
| eagle          | rap           | doca                     |
| e-book         | rash          | dolar                    |
| ecobag         | rave          | drawback                 |
| ecstasy        | ray-ban       | drible                   |
| e-mail         | ready-made    | drink                    |
| endomarketing  | reality show  | drive-in                 |
| enlightment    | recall        | drop                     |
| enter          | referee       | drugstore                |
| escort         | reggae        | esporte                  |
| ethernet       | relax         | estêncil                 |
| exabyte        | release       | estoque                  |
| face-lift      | relish        | expert                   |
| face-lifting   | remake        | fã                       |
| factoring      | replay        | faroeste                 |
| fade           | replicon      | fast-back                |
| fade-in        | reprint       | filme                    |
| fair-play      | resort        | flanela                  |
|                |               | flash                    |

|                  |                         |              |
|------------------|-------------------------|--------------|
| fairway          | resurfacing             | flerte       |
| fanzine          | revival                 | folclore     |
| fashion          | rhythm and blues        | footing      |
| fast-food        | rock                    | foxtrot      |
| fax-modem        | rock and roll           | freeway      |
| feature          | round                   | front        |
| feedback         | royal straight<br>flush | futebol      |
| ferry            | royalty                 | galão        |
| fifo             | rubber                  | gang         |
| finger           | rush                    | garden-party |
| firewall         | saloon                  | gateway      |
| fitness          | scanner                 | gilete       |
| five             | scholar                 | gim          |
| five o'clock tea | scraper                 | gol          |
| flap             | script                  | golfe        |
| flash            | sedan                   | grapefruit   |
| flashback        | self-made man           | hall         |
| flat             | self-service            | happening    |
| flint-glass      | serial killer           | happy-end    |
| float            | set                     | hit parade   |
| flip-flop        | set-point               | hobby        |
| floater          | setup                   | holding      |
| flush            | sex appeal              | hóquei       |
| flutter          | sex shop                | hostess      |
| flyback          | shadow price            | iceberg      |
| fog              | shape                   | jarda        |
| folder           | share-of-market         | jeep         |
| follow-up        | share-of-mind           | jazz         |
| footing          | shareware               | jersey       |
| forehand         | shimmy                  | jingle       |
| forward          | shiraz                  | jóquei       |
| foul             | shopping                | júri         |
| four             | shopping center         | ketchup      |
| fox-blue         | short                   | leiaute      |
| fox-terrier      | short                   | líder        |
| franchise        | show                    | long play    |
| frankenstein     | showroom                | lorde        |
| freelance        | shunt                   | lunch        |
| freelancer       | sib                     | lanchonete   |
| free shop        | sibling                 | magazine     |
| freeware         | side-car                | marketing    |
| freezer          | silk screen             | mass media   |
| front            | sitcom                  | match        |
| frontside        | site                    | miss         |
| fuel             | skate                   | motel        |
| full-back        | sketch                  | napalm       |
| full-contact     | skinhead                | nocaute      |
| full hand        | sky-surf                | office-boy   |
| full time        | slack                   | offset       |

|               |                   |                 |
|---------------|-------------------|-----------------|
| funding       | slice             | one-man show    |
| funding loan  | slide             | orlon           |
| funk          | slip              | outdoor         |
| fuzzy         | slogan            | pênalti         |
| fuzzy logic   | smarthphone       | performance     |
| gag           | smash             | pick-up         |
| game          | smoking           | pickles         |
| gang          | snowboard         | pingue-pongue   |
| gangster      | soap opera        | piqueenique     |
| gatekeeper    | soccer            | photochart      |
| gateway       | socialite         | place           |
| ghost-writer  | softball          | playboy         |
| gentleman     | soft news         | playground      |
| gigabit       | software          | pôquer          |
| gigabyte      | songbook          | raid            |
| ginger ale    | soul              | raquete         |
| ginseng       | spaghetti western | recital         |
| girl          | spam              | recorde         |
| glamor        | spammer           | refill          |
| glamour       | sparing           | repórter        |
| glide         | speaker           | revolver        |
| globe-trotter | speech            | rifle           |
| gloss         | spencer           | ringue          |
| go-go boy     | spin              | road show       |
| golme         | spinning          | rosbife         |
| good bye      | spot              | rounde          |
| gospel        | spray             | rum             |
| gossip        | sprinkler         | rush            |
| grade         | sprinter          | sanduíche       |
| grapefruit    | spyware           | science fiction |
| gray          | squas             | script          |
| green         | squid             | sedã            |
| grid          | staff             | shampoo         |
| grill-room    | stand             | shorts          |
| grip          | standard          | show            |
| groom         | stand-by credit   | sketch          |
| grunge        | star              | slide           |
| habook        | starter           | slogan          |
| hacker        | starting gate     | smoking         |
| hakka         | states            | society         |
| half-back     | steeple-chase     | soccer          |
| hall          | step              | sonar           |
| hamster       | still             | speaker         |
| handball      | stop              | staff           |
| handicap      | storyboard        | estande         |
| hangfive      | straight flush    | starlet         |
| hangten       | stress            | steward         |
| happening     | strip-tease       | stroke          |
| happy ending  | strip-teaser      | stud            |
| happy few     | stud              | suéter          |

|               |                  |          |
|---------------|------------------|----------|
| happy hour    | sucker           | surf     |
| hard news     | sundae           | tape     |
| hard rock     | supplier' credit | teenager |
| hardware      | surf             | tênder   |
| headhunter    | suspect          | tênis    |
| heavy-metal   | swab             | taeste   |
| hedge         | swarmers         | time     |
| help          | sweepstake       | trailer  |
| high fidelity | swell            | tramp    |
| high tech     | swing            | truste   |
| hip hop       | syrah            | túnel    |
| hippie        | tablet           | turfe    |
| hit           | take             | twist    |
| hit man       | talk-show        | uísque   |
| hit parade    | tape deck        | voleibol |
| hobby         | tape recorder    | warrant  |
|               |                  | watt     |
|               |                  | weekend  |

#### **D) Termo de Consentimento Livre Esclarecido do experimento psicolinguístico**

Universidade Federal do Rio Grande do Sul  
Programa de Pós-graduação em Letras  
Estudos da Linguagem: Fonologia e Morfologia

O(A) Sr.(a) está sendo convidado(a) a participar do projeto de pesquisa intitulado Experimento psicolinguístico online para o teste de pistas formais na atribuição de gênero gramatical em português brasileiro, sob a responsabilidade do pesquisador e orientador Dr. Luiz Carlos Schwindt e condução do pesquisador Pedro Surreaux, do curso de pós-graduação em Letras. Para participar, você precisa ter mais que 18 anos e ser falante nativo do português brasileiro.

Esta pesquisa tem o objetivo de entender os fatores envolvidos na percepção do gênero gramatical na estrutura dos substantivos em português. O projeto foi avaliado pelo CEP-UFRGS, órgão colegiado, de caráter consultivo, deliberativo e educativo, cuja finalidade é avaliar, emitir parecer e acompanhar os projetos de pesquisa envolvendo seres humanos, em seus aspectos éticos e metodológicos, realizados no âmbito da instituição. Tal órgão pode ser contato pelo telefone (51) 3308-3787.

Para a instituição e para a sociedade, esta pesquisa servirá como parâmetro para avaliar as variáveis envolvidas no processo de pluralização de nomes e adjetivos no português brasileiro. Em torno de 230 participantes irão responder a este questionário. Os dados coletados, serão, posteriormente, discutidos e analisados, o que contribuirá para a literatura especializada da área da linguística. Todos os participantes serão maiores de idade e terão como língua materna o

português brasileiro. Além disso, para participar é preciso ter acesso estável a rede de internet e a um dispositivo eletrônico que você possa usar para realizar este questionário (telefone, celular, tablet, computador).

A pesquisa é feita de forma voluntária e não apresenta remuneração ou algum benefício direto. Indiretamente, um possível benefício é a reflexão sobre o uso da língua falada e da língua escrita no dia a dia.

Quanto aos riscos, a participação na pesquisa não deve ocasionar nenhum dano moral aos participantes; no entanto, alguns participantes podem apresentar desconforto por olhar continuamente para a tela do computador ou celular, ou cansaço mental e ocular ao decorrer do experimento. Esses riscos, caso se concretizem, serão de caráter passageiro, e não permanente. A forma de minimizá-los será com a elaboração de um experimento curto e com frases de fácil leitura. Em virtude de as informações coletadas serem utilizadas unicamente com fins científicos, garantimos o total sigilo e confidencialidade, por meio da concordância com este termo. Para concordar, você deve clicar em Eu compreendo as condições desta pesquisa. Além disso, o(a) Sr.(a) terá direito e a liberdade de negar-se a participar desta pesquisa total ou parcialmente ou dela retirar-se a qualquer momento, sem que isto lhe traga qualquer prejuízo, de acordo com a Resolução CNS 510/2016 e complementares. Caso você queira conhecer os resultados desta pesquisa, estes ficarão à disposição com os responsáveis.

Serão solicitadas informações como sua idade e sua escolaridade; entretanto, elas não possuem o objetivo de identificar o participante. O presente experimento possui caráter anônimo, portanto. Todas as informações coletadas nesta pesquisa são estritamente confidenciais. Trataremos todas as informações sem que haja identificação de particularidades de cada participante. Caso sinta algum tipo de desconforto, é possível parar a sua participação a qualquer momento e fazer uma pausa para retornar ao questionário. Em seguida, será solicitado seu consentimento em participar da pesquisa. A assinatura deste Termo não exclui a sua possibilidade de buscar indenização diante de eventuais danos decorrentes de participação na pesquisa, como preconiza a Resolução 466/12. O participante, também, terá acesso ao registro de consentimento sempre que solicitá-lo.

Para qualquer esclarecimento no decorrer da sua participação, favor contatar o pesquisador responsável ([schwindt@ufrgs.br](mailto:schwindt@ufrgs.br)) ou o pesquisador assistente ([surreauxpp@gmail.com](mailto:surreauxpp@gmail.com)) por e-mail ou no endereço profissional, na Avenida Bento Gonçalves, 9500 – Agronomia – Porto Alegre/RS – Instituto de Letras, sala 212, ou o Comitê de Ética em Pesquisa da universidade pelo telefone (51)33083738. Desde já, agradecemos.

Pesquisador responsável  
Luiz Carlos da Silva Schwindt  
[schwindt@ufrgs.br](mailto:schwindt@ufrgs.br)

Pesquisador assistente  
Pedro Surreaux



surreauxpp@gmail.com

## **E) Materiais do questionário**

Palavras do experimento

Palavras-teste (45): vação, lareção, bolarição, raca, taluza, falinara ruzão, calestão, valenião, uzência, nafiquência, tolicanência, landade, matidade, garovidade, vacinéi, opiniéi, televiséi, cidadêr, inspiracêr, imaginacêr, coisive, historive, bicicletive, noitofe, laranjofe, brincadeirofe, bole, semane, , almofade, dezéi, crolinéi dostamenéi, racêr, voluzêr, mencalinêr, potive, lerobive, evarodive, unofe, evizofe, nofilarofe, sove, dersuve, renulique.

Palavras distratoras (15): científico, verdura, candidato, trânsito, mão-de-obra, óleo, grupinho, galinho, gaita, durabilidade, desordeiro, brinquedo, vizinho, tempero, aeronáutica.

Frase-veículo

Frase 1: Já sabemos muito sobre \_\_ [palavra] .

Alternativas:

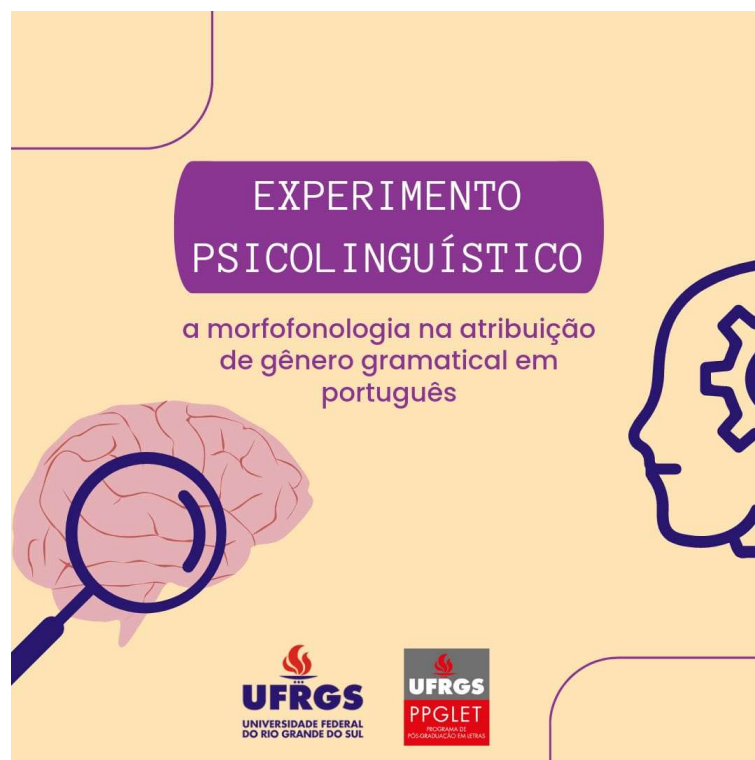
Conjunto 1:

- a) a
- b) o

Conjunto 2:

- a) o
- b) a

## **F) Convite para a participação no experimento psicolinguístico**



Se você é falante nativo de português brasileiro e tem mais de 18 anos, te convido a colaborar com minha pesquisa de mestrado sobre o sistema de gênero gramatical da nossa língua. Para isso, clique no link e complete o questionário, que dura em média 5 minutos para ser respondido. Nele, você encontrará algumas palavras que já conhece e outras que foram elaboradas para este estudo, e deverá escolher o artigo definido – a ou o, feminino ou masculino, – que soa mais natural em cada uma das frases.

Agradeço muito pela sua participação e, se possível, compartilhe com outras pessoas!