

FALSOS DIMINUTIVOS DO PORTUGUÊS BRASILEIRO E SEU RECONHECIMENTO EM UM DICIONÁRIO COMPUTACIONAL DE LIVRE ACESSO

FALSE DIMINUTIVES OF BRAZILIAN PORTUGUESE AND THEIR RECOGNITION IN A FREELY AVAILABLE COMPUTATIONAL DICTIONARY

Roana Rodrigues (UFS)¹
Oto Araújo Vale (UFSCar)²

Resumo: Neste trabalho propõe-se a criação de uma lista de falsos diminutivos do português brasileiro (PB) e a análise e descrição de sua cobertura por um dicionário computacional (DELAF-BR), que, assim como outros dicionários (léxicos) computacionais, é um recurso basilar para o processamento da linguagem e depende da descrição minuciosa e sistemática da língua. Para tanto, analisaram-se 953 lexemas terminados em *-inho/-inha*, retirados de dicionários e de *corpus* do PB, classificando-os em: (i) diminutivos reais (Dim), podendo designar grau de tamanho (*botinha*), intensidade (*cedinho*), juventude (*homenzinho*) e juízo de valor (*povinho*); e (ii) falsos diminutivos (FDim), em que o significado do lexema se distancia da palavra da qual deriva (*almofadinha*, *camisinha*, *carrinho*). Com a lista de 468 FDim construída, foi analisado o seu processamento pelo DELAF-BR e verificou-se que apenas 31 lexemas descritos (6,6%) constavam nesse dicionário eletrônico com entradas independentes. Considerando-se ainda os 37 FDim mais frequentes em *corpora* de língua geral, o resultado foi semelhante: apenas 10,8% dos casos (4 lexemas) estavam contemplados no dicionário. Sendo assim, esta pesquisa apresenta contribuições não só para os estudos descritivos da língua, mas, sobretudo, para a área de Processamento de Língua Natural com a construção de um recurso lexical de FDim do PB.

Palavras-chave: Recurso lexical; Sufixação; Falso diminutivo; Processamento de língua natural.

Abstract: In this paper we propose the creation of a list of false diminutives in Brazilian Portuguese (BP) and the analysis and description of their coverage by a computational dictionary (DELAF-BR), which, like other computational dictionaries (lexicons), is a basic resource for language processing and depends on the meticulous and systematic description of the language. To this end, we analyzed 953 lexemes ending in *-inho/-inha* and we classified them into: (i) real diminutives (Dim), which can designate degree of size (*botinha*: little boot), intensity (*cedinho*: early in the morning), youth (*homenzinho*: young boy) and value judgment (*povinho*: ordinary people); and (ii) false diminutives (FDim), in which the meaning of the lexeme distances itself from the word from which it derives (*almofadinha* - prig, *camisinha* - condom, *carrinho* - sliding tackle). With the list of 468 FDim built, we analyzed its processing by DELAF-BR and we verified that only 31 described lexemes (6.6%) were included in this electronic dictionary. Considering the 37 most frequent FDim in general language corpora, the result was similar: only 10.8% of the cases (4 lexemes) were contemplated in the dictionary. Therefore, this research presents contributions to descriptive

¹ Professora do Departamento de Letras Estrangeiras (DLES) e do Programa de Pós-Graduação em Letras (PPGL) da Universidade Federal de Sergipe. <https://orcid.org/0000-0002-7748-8716>. E-mail: roana@academico.ufs.br

² Professor do Departamento de Letras (DL) e do Programa de Pós-Graduação em Linguística (PPGL) da Universidade Federal de São Carlos (UFSCar). <https://orcid.org/0000-0002-0091-8079>. E-mail: otovale@ufscar.br

studies of language and to the area of Natural Language Processing with the construction of a lexical resource of the false diminutives in Brazilian Portuguese.

Keywords: Lexical Resource; Suffixation; False diminutive; Natural language processing.

Introdução³

O estudo dos diminutivos (*camiseta: camisetinha*) e falsos diminutivos (*camisa ≠ camisinha*) do PB é um fenômeno bastante conhecido e descrito sob diferentes bases teórico-metodológicas (ALVES, 2006; MORAES, OLIVEIRA, 2010; MENDES, 2012; RODRIGUES, VALE, 2013⁴). No entanto, esta pesquisa insiste na necessidade de uma retomada dos estudos dos diminutivos e falsos diminutivos a partir de uma nova perspectiva: a da construção de um recurso léxico – criação de uma lista de falsos diminutivos – e de sua contribuição direta para a área de Processamento de Língua Natural (PLN). Segundo Ferreira e Lopes (2021, p. 10), o PLN, também chamado de Linguística Computacional (LC), é estudado por linguistas e cientistas da computação, com o intuito de “criar máquinas capazes de compreender e produzir expressões em língua humana”.

Em trabalho anterior, Rodrigues, Souza e Santos (2022) enfatizam a via de mão dupla entre a *descrição linguística* e a *computação*, a partir da necessidade de trabalhos descritivos linguísticos para a criação de dados seguros para o processamento da máquina, e da demanda por recursos computacionais de ponta para a utilização em pesquisas e ensino. Sendo assim, estudos descritivos minuciosos e sistemáticos sobre os mais variados fenômenos linguísticos são fundamentais para a aplicação e desenvolvimento de sistemas. Um recurso essencial para o PLN são os dicionários computacionais, que funcionam como partes de programas específicos em sistemas informatizados. Segundo Finatto *et al.* (2019, p. 64), esses dicionários, também conhecidos como *léxicos computacionais*, são usados para o processamento da linguagem e, para o seu bom desenvolvimento, dependem de uma série de informações linguísticas descritas de maneira cuidadosa.

Assim como no trabalho de Finatto *et al.* (2019), na presente investigação será analisado, dentre outros fatores, como o dicionário computacional DELAF-PB (MUNIZ, 2004; CALCIA *et al.*, 2014; CALCIA, 2015; VALE, BAPTISTA, 2015) do sistema Unitex (PAUMIER, 2021) processa morfologicamente os casos de falsos diminutivos estudados nesta pesquisa. Salienta-se a importância na seleção desse dicionário específico, já que ele serve de base para o desenvolvimento de outros recursos léxico-computacionais, como é o caso do projeto PortiLexicon-UD⁵.

Sendo assim, este trabalho tem como objetivo principal a criação de uma lista de lexemas terminados em *-inho/-inha* que se enquadram como falsos diminutivos (diminutivos não

³ Alguns resultados desta investigação foram apresentados na 29ª Jornada Internacional do GELNE (Grupos de Estudos Linguísticos e Literários do Nordeste), no simpósio temático “Aplicações e desafios do uso de ferramentas computacionais na linguística”. Essa edição da Jornada promoveu discussões linguísticas e literárias sobre as mais variadas e atuais temáticas, dentre as quais destacam-se os debates relacionados às tecnologias, tão importantes para a continuidade do trabalho em meio à pandemia, tanto no que diz respeito à sua aplicação ao ensino e pesquisa, quanto à mobilização de recursos linguísticos para aplicações computacionais.

⁴ Apesar de a lista de diminutivos e falsos diminutivos já ter sido apresentada em um trabalho anterior (RODRIGUES, VALE, 2013), este artigo consolida aqueles dados após revisão e ampliação, além da inclusão da análise de seu reconhecimento em um dicionário computacional.

⁵ O projeto PortiLexicon-UD (a lexicon for Brazilian Portuguese according to Universal Dependencies) contém mais de 1 milhão de lemas do português com suas informações morfológicas e morfossintáticas, seguindo o modelo da Universal Dependencies (UD). O léxico se baseia, entre outros sistemas, no Unitex-PB, que é um *software* de processamento de textos a partir da utilização de dicionários eletrônicos em formato DELAF (LOPES *et al.*, 2022). Disponível em: <https://portilexicon.icmc.usp.br/>. Acesso em setembro de 2022.

composicionais), pois o seu significado se distancia do significado do nome do qual deriva (*caninha, lanterninha, quentinha, tampinha*); e a avaliação da cobertura dessa lista pelo dicionário computacional DELAF-BR. Já, como objetivos específicos, a pesquisa possibilita a discussão e sistematização dos lexemas terminados em *-inho/-inha*, apresentando as nuances de classificação dos diminutivos reais (diminutivos composicionais), para além dos casos de falsos diminutivos.

De fato, o elemento que motivou inicialmente esta pesquisa foi uma observação feita⁶ a partir de Silva *et al.* (2010) de que a lexia *camisinha* aparecia sempre normalizada como *camisa*, o que dificultava a extração de informação em textos sobre a prevenção de doenças sexualmente transmissíveis. Portanto, com este trabalho, espera-se discutir as seguintes questões:

- Qual o comportamento e proposta de classificação dos lexemas terminados em *-inho/-inha* do PB?
- Como um dicionário computacional processa os lexemas identificados como falsos diminutivos do PB?

Metodologicamente, este trabalho, de base quantitativa e qualitativa, realizou-se em quatro etapas: (i) extração, análise e descrição de lexemas terminados em *-inho/-inha* em dois dicionários de língua portuguesa: Houaiss (HOUAISS, 2009) e Aurélio (FERREIRA, 2009); e (ii) extração, análise e descrição das primeiras 1.000 palavras mais frequentes com tais terminações no *corpus* do NILC/São Carlos (constituído por textos jornalísticos, didáticos, epistolares e redações de alunos)⁷, a fim de constatar a possibilidade de acréscimo de lexemas terminados em *-inho/-inha* e observar o comportamento dessas palavras em uso. Em um segundo momento, através do *software* Unitex (PAUMIER, 2021), realizou-se: (iii) verificação da frequência dos falsos diminutivos do PB em dois *corpora* de textos jornalísticos (Folha-94-02 e Folha Kaggle); e (iv) análise da cobertura dos falsos diminutivos pelo dicionário computacional DELAF-BR, com o objetivo de analisar a necessidade de inclusão desses dados para o aperfeiçoamento lexical de recursos que podem ser utilizados na área de PLN.

Para a descrição detalhada dos dados coletados e analisados, este artigo se organiza da seguinte maneira: na primeira seção, são retomados conceitos e reflexões a respeito dos sufixos *-inho/-inha* do PB. Em seguida, são apresentados os processos metodológicos e a análise e classificação propostas para os casos de diminutivos e falsos diminutivos, com base nos dados retirados dos dicionários e do *corpus* do NILC. Na terceira seção, tem-se a descrição do dicionário computacional analisado (DELAF-PB), considerando, através do uso do *software* Unitex (PAUMIER, 2021), a frequência dos lexemas anotados como falsos diminutivos em *corpora* de língua geral e a verificação da cobertura de tais casos pelo dicionário eletrônico analisado. Por fim, apresentam-se as conclusões e contribuições gerais desta investigação.

1 Reflexões a respeito dos sufixos *-inho/-inha* do PB

Em seu trabalho didático sobre os estudos morfológicos, Petter ([2003] 2021, p. 62) descreve, sucintamente, a complexidade inerente à morfologia, área que se dedica “ao estudo da forma das palavras”, visto que a própria noção de *palavra* é bastante complicada e controversa. A autora descreve o funcionamento de alguns processos morfológicos e apresenta a subdivisão convencional da área em dois grandes campos: morfologia flexional e morfologia lexical, em que o primeiro se dedica, *grasso modo*, à análise dos mecanismos morfológicos que apresentam

⁶ Júnia Anacleto em comunicação pessoal, 2010.

⁷ Corpus do NILC/São Carlos, disponível na Linguateca, em: <https://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>. Acesso em setembro de 2022.

informações gramaticais, cujo procedimento básico é a flexão (*cont-ei, cont-ou*); e o segundo se refere aos mecanismos morfológicos por meio dos quais se formam novas palavras, sobretudo pelos processos de *derivação* (*legal > i-legal; gato > gat-inho*) e de *composição* (*sofá+cama > sofá-cama*).⁸

Apesar de os afixos de grau (diminutivo, aumentativo, intensidade) já terem sido categorizados como flexionais, a partir da década de 1970, principalmente com as contribuições de Mattoso Câmara Jr., esses afixos, que englobam os sufixos -inho/-inha, objetos desta pesquisa, passam a ser classificados como parte dos estudos da *morfologia lexical*, relacionados, mais especificamente, aos processos de *derivação*. Segundo Gonçalves (2019, p. 153), apesar de que “a maior parte dos livros didáticos destinados ao ensino médio ainda insist[a] em colocar os afixos de grau no mesmo patamar do de gênero e do de número”, na atualidade, nenhum teórico – nem em gramáticas normativas – considera a gradação morfológica um processo flexional.

A fim de compreender as particularidades dos processos de derivação morfológica de grau, Gonçalves (2019) propõe alguns critérios formais para a sua identificação, enfatizando, no entanto, que tais critérios devem ser vistos como um recurso didático e não um veredito final acerca desses afixos, a saber: (i) isenção do requisito de “obrigatoriedade sintática”, ou seja, é facultativa a utilização dos afixos gradativos na construção do sintagma nominal, como em *carrinho bonitinho > carro bonitinho > carrinho bonito*; (ii) vinculação de um significado que pode ser parafraseado por outra forma de expressão: *carrinho > carro pequeno*; (iii) produtividade de uso, pois esses sufixos podem anexar-se a nomes (*carrinho*), pronomes (*euzinho*), advérbios (*pertinho*), interjeições (*adeusinho*) e verbos (*correndinho*); (iv) promoção de alterações de categoria lexical: *branco* (adjetivo) > *branquinho* (substantivo); (v) variabilidade semântica, pois o significado pode variar de uma palavra para outra: -inho/-inha pode expressar dimensão, apreço, despreço ou afeto; (vi) pouca restrição quanto ao número de afixos que admite uma palavra: *livrãozinho, vidinhazinha*; e (vii) arbitrariedades e desvios de maneira frequente, como os usos figurados de sufixos aumentativos e diminutivos: *coxinha* (salgado), *selinho* (beijo) e *bolão* (aposta).

Em gramáticas tradicionais, como as de Cunha e Cintra (2008), Azeredo (2008) e Bechara (2009), os sufixos -inho/-inha e -ão/-ona⁹ são definidos como aqueles mais frequentes em língua portuguesa para a expressão da variação de grau, podendo adicionar uma referência de dimensão física (tamanho diminuído ou tamanho aumentado), além de denotar juízo de valor (depreciativo, apreciativo ou afetivo) ao lexema.

Nos estudos descritivos da língua, os sufixos -inho/-inha são analisados de maneira aprofundada, a partir de diferentes perspectivas teórico-metodológicas, considerando a sua complexidade e usos reais. Alves (2006), em sua investigação de cunho funcional-discursivo, por exemplo, atribui três valores para o sufixo -inho: *tamanho/quantidade/intensidade reduzida, afetividade e avaliação positiva ou negativa*, ressaltando a importância da atribuição do sufixo como uma estratégia comunicativa e considerando fatores extralinguísticos para a sua interpretação. Por sua vez, o estudo pragmático de Moraes e Oliveira (2010), com a análise de um programa de televisão, aponta para a classificação dos usos dos lexemas diminutivos em sete categorias: *afetividade, ênfase, ironia, tamanho reduzido, uso pejorativo, eufemismo e significação própria sem ligação com a palavra de origem*.

Verifica-se, portanto, que os sufixos -inho/-inha se associam à noção de grau de tamanho e aos valores de afetividade e avaliação (positiva ou negativa), essas últimas, considerando fatores extralinguísticos, como os interlocutores e a situação comunicativa. Além disso, um aspecto mencionado em alguns desses estudos (MORAES, OLIVEIRA, 2010; GONÇALVES, 2019) é a

⁸ Embora se convença organizar os estudos em *morfologia flexional* e *morfologia lexical*, há investigações, como as de Gonçalves (2019, 2020), que apresentam as diferenças, semelhanças e, inclusive, um *continuum* entre os processos de flexão e derivação.

⁹ Neste artigo, são apresentados os dados referentes apenas aos casos de diminutivos reais e falsos diminutivos do PB. No entanto, está em fase de revisão um estudo semelhante referente ao comportamento dos sufixos -ão/-ona, relacionados às noções de aumentativos reais e falsos aumentativos.

existência de usos figurados desses sufixos, caracterizados por apresentarem uma significação própria, que são os casos dos lexemas terminados em -inho/-inha como *camisinha* e *calcinha*, que se distanciam do significado de suas palavras de origem (*camisa* e *calça*).

Sobre esse último caso em particular, pode-se citar as contribuições de Armelin (2014), que classifica os sufixos -(z)inho e -(z)ão¹⁰ como *composicionais* e *não composicionais*. A interpretação *composicional* se refere ao que nesta pesquisa se denomina *diminutivos reais* (*carrinho*: carro pequeno), denotando o grau de tamanho ou os valores de afetividade e avaliação do nome do qual deriva (nome original); já a interpretação *não composicional* é aqui entendida como *falsos diminutivos*, designando um distanciamento semântico entre o nome derivado e o nome original (*caipirinha*: bebida alcoólica).

Segundo Armelin (2014, p. 399-400), diminutivos e aumentativos não composicionais parecem ser capazes de determinar propriedades formais da estrutura. A título de exemplo, a autora menciona casos em que substantivos não animados e de gênero feminino, como *almofada*, podem se tornar animados e ter variação de gênero a depender do referente (*o/a almofadinha*: pessoa mimada [sic]; pessoa que se veste com apuro exagerado). Além disso, pode-se verificar a ocorrência de alterações na classe gramatical do nome original do diminutivo não composicional: *quente* (adjetivo) > *quentinha* (substantivo). Armelin (2014, p. 399) ainda pontua que algumas dessas formas podem ter interpretações ambíguas, considerando seu uso composicional (*almofadinha*: almofada pequena) e não composicional (*almofadinha*: pessoa mimada [sic]).

Essas pesquisas demonstram a produtividade desses sufixos no PB, assim como a abrangência e importância de seu estudo. Considerando os objetivos e as especificidades da presente pesquisa, na próxima seção, serão apresentadas as orientações metodológicas utilizadas e a descrição e classificação dos lexemas terminados em -inho/-inha coletados nesta investigação.

2 Obtenção, análise e classificação dos lexemas terminados em -inho/-inho do PB

Nesta pesquisa, a obtenção da lista de lexemas terminados em -inho/-inha ocorreu em dois momentos. Primeiramente, foram analisadas as entradas com tais terminações em dois dicionários (em formato eletrônico – CD-ROM) de língua portuguesa: Houaiss (HOUAISS, 2009) e Aurélio (FERREIRA, 2009).¹¹ Em um segundo momento, foram investigadas as 1.000 (mil) primeiras entradas terminadas em -inho e também em -inha no *corpus* do NILC/São Carlos, a partir de buscas realizadas no site da Linguateca, objetivando verificar a possibilidade de novas palavras a serem incluídas na lista de falsos diminutivos.

A escolha dos dicionários como fonte inicial de informação deveu-se à necessidade de um estudo quantitativamente relevante de tais sufixos. A Tabela 1 apresenta o número de lexemas encontrados.

¹⁰ Para Armelin (2014, p. 400), as formas -inho/-ão são passíveis de serem não composicionais, enquanto -zinho/-zão são composicionais, ou seja, se referem a formas diminutivas e aumentativas reais. A autora salienta, no entanto, que -zinho/-zão podem compor uma construção não composicional quando a consoante /z/ é exigida por uma restrição fonológica imposta pela língua: café > cafeinho* > cafezinho, cujo significado composicional é *café pequeno*, e o significado não composicional se refere a uma refeição informal e pequena.

¹¹ Nos CD-ROM de cada dicionário, foi possível fazer buscas por lexemas terminados em -inho e em -inha. Os resultados gerados foram copiados manualmente em planilhas para posterior análise.

Tabela 1 – Dados numéricos dos lexemas terminados em -inho/-inha extraídos dos dicionários

Terminação	Quantidade
-inho	477
-inha	465
Total	942 ¹²

Fonte: Elaboração própria

Com a lista estabelecida, cada lexema foi analisado individualmente de maneira manual, considerando as informações disponíveis na microestrutura dos dicionários (cabeça do verbete, definições e exemplos). Os dados foram dispostos em tabelas binárias, em que as linhas contêm a entrada lexical e as colunas, as informações anotadas. Quando determinada entrada apresenta a propriedade em questão, é assinalado o símbolo “+”, quando o lexema não possui a propriedade, aponta-se o símbolo “-”. O Quadro 1 exemplifica a disposição e organização dos dados.

Quadro 1 – Propriedades dos lexemas terminados em -inho/-inha extraídos dos dicionários¹³

Dicio		Propriedades do lexema							Campo semântico					Tipo de diminutivo			
Aurélio	Houaiss	Lema	FOrig	NComp	FOrig	Dim	FDim	Sin	Botânica	Ornitologia	Ictiologia	Aguardente	Esporte	Tamanho	Intensidade	Juventude	Juízo de valor ¹⁴
+	+	ancinho	ancinho	-	+	-	-	-	-	-	-	-	-	-	-	-	-
+	+	maminha	mama	-	-	+	+	-	-	-	-	-	-	+	-	-	-
+	+	povinho	povo	-	-	+	-	-	-	-	-	-	-	-	-	-	+
+	+	purinha	pura	-	-	+	+	-	-	-	-	+	-	-	+	-	-
+	+	cestinha	cesta	-	-	+	+	-	-	-	-	-	+	+	-	-	-

Fonte: Elaboração própria

No Quadro 1, é possível observar a existência de diferentes lexemas com a terminação -inho/-inha, considerando: (i) lexemas ambíguos, que podem atuar tanto como diminutivos reais, quanto como falsos diminutivos: *maminha* (mama pequena ou corte de carne bovina); *cestinha* (cesta pequena ou termo do basquete que se refere ao jogador que faz o maior número de pontos); *purinha* (muito pura ou sinonímia de cachaça); (ii) caso de lexema *pseudossufixado*: *ancinho* (ferramenta agrícola, cuja terminação -inho não se refere ao sufixo diminutivo estudado); e (iii) diminutivos reais: *povinho* (juízo de valor à *povo*).

A partir da lista estabelecida, foi possível organizar os dados da seguinte maneira: lexemas desconsiderados na pesquisa (318 casos); lexemas diminutivos reais (535 casos); e lexemas falsos

¹² Das 942 entradas de lexemas terminados em -inho/-inha, 790 estão nos dois dicionários, em que: 72 entradas estão contempladas apenas no Aurélio (2009) e 80, apenas no Houaiss (2009).

¹³ Notação: Dicio: dicionários; Lema: entrada do dicionário; FOrig: forma original da palavra, ou seja, a forma primitiva da qual deriva o lema; Ncomp: nome composto; Dim: diminutivo real (cuja especificidade consta no campo “tipo de diminutivo”); FDim: falso diminutivo; Sin: sinonímia, pois a definição no dicionário apresenta a expressão “mesmo que” e faz remissão à forma original do lexema; Campo semântico: áreas de conhecimento que, no momento da anotação, apresentaram-se como relevantes por sua frequência; Tamanho: diminutivo real por grau de tamanho (pequeno); Intensidade: diminutivo real grau de intensidade (muito, pouco); diminutivo por designação de juventude (jovem x); diminutivo por juízo de valor (efetividade, ironia, avaliação positiva ou avaliação negativa).

¹⁴ Nesta pesquisa, *juízo de valor* refere-se ao uso do sufixo diminutivo para incluir ao lexema a informação de uma avaliação positiva, negativa ou ainda para designar uma forma de tratamento afetivo.

diminutivos (457 casos); tendo havido 368 casos anotados tanto como diminutivos reais, quanto como falsos diminutivos.

Lexemas desconsiderados na pesquisa foram aqueles que se constituem por nomes compostos, tais como *porco-espinho*, *abóbora-moranguinha* (135 casos) e/ou foram anotados como *pseudossufixos*, ou seja, casos nos quais as terminações não apresentam relação com os afixos diminutivos estudados (183 casos), como *adivinbo*, *cominbo*, *cozinba*. Segundo o Grupo de Morfologia Histórica do Português, GMHP (2014, p. 13), “muitas palavras com uma *terminação* idêntica não contêm necessariamente o mesmo *sufixo*”. A partir de uma perspectiva diacrônica, o Grupo exemplifica o fenômeno de lexemas pseudossufixados, também denominados *falsos sufixos*, como o caso de *macaxeira*, palavra derivada do tupi, cuja terminação *-eira* não equivale ao sufixo encontrado em *pereira*, palavra que provém do latim. Segundo os autores, é uma etapa necessária e inicial o descarte de palavras pseudossufixadas no estudo morfológico. Sendo assim, do total de 942 entradas lexicais, foram consideradas para fins classificatórios 624 casos.

Desse valor total (624 lexemas), 535 foram anotados como diminutivos reais e, assim como já estabelecido em gramáticas e trabalhos descritivos anteriores, o diminutivo real engloba não apenas a noção de grau de tamanho (*brinquinbo*: brinco pequeno; *boizinho*: boi pequeno), como também o grau de intensidade (*devagarinbo*: muito devagar; *chuvinha*: pouca chuva, chuva fraca), a designação de juventude (*bomenzinho*: homem jovem) e a apresentação juízo de valor apreciativo, depreciativo ou de afetividade (*amorzinho*, *artinba*, *filbinbo*, *povinho*).

Ressaltam-se ainda casos em que a forma diminutiva foi classificada como *sinonímia* da palavra de origem, tais como *salsinha* = *salsa*; *queijadinha* = *queijada*. Embora tenham sido anotados os lexemas cujas definições nos dicionários faziam uma remissão a outro lexema sem a terminação *-inho/-inha*, antecedida da expressão “*mesmo que*”, optou-se, neste trabalho, por relacionar esses casos às formas diminutivas reais, pois, como aponta Alves (2006, p. 694), apesar de os recursos morfológicos se aplicarem ao nível da palavra, eles são “alvo da expressão de realidades pragmáticas e discursivas”, sendo comumente utilizados como estratégia de comunicação para designar afetividade ou avaliação (positiva ou negativa).

Por fim, 457 entradas foram classificadas como falsos diminutivos por se distanciarem das noções de grau de tamanho, grau de intensidade, designação de juventude e/ou juízo de valor. Trata-se, portanto, dos casos de diminutivos não composicionais (ARMELIN, 2014), que apresentam uma significação específica que, talvez por um processo metafórico, afastou-se do significado da palavra da qual deriva, como ocorre nos seguintes casos: *covinba*: pequena cavidade no queixo; *flanelinha*: pessoa que vigia veículos nas ruas; *santinho*: panfleto com foto e identificação de um candidato, distribuído em época de eleições; *vaquinha*: coleta de dinheiro por um grupo de pessoas para o pagamento de uma despesa comum.

Salienta-se que 369 lexemas foram anotados como diminutivo real e falso diminutivo, ou seja, entradas que permitem tanto a interpretação composicional (*beijinho*: pequeno beijo), quanto a interpretação não composicional (*beijinho*: doce tipicamente brasileiro de leite condensado, coco e cravo, servido em festas).

Sabe-se que tanto os casos de diminutivos reais, quanto os casos de falsos diminutivos, apresentam variados significados, no entanto, durante o processo de anotação, foram encontrados os seguintes campos semânticos recorrentes com os sufixos *-inho/-inha* investigados: botânica, com 123 lexemas (*baratinha*, *jangadinha*); ornitologia, com 78 lexemas (*chapinha*, *tucaninho*); ictiologia, com 44 lexemas (*noivinha*, *tesourinha*); sinonímia de aguardente, com 21 lexemas, todos terminados em *-inha* (*azulzinha*, *caninha*), e lexemas relacionados ao esporte, com 8 casos (*bandeirinha*, *carrinho*, *cestinha*, *chuveirinho*).

Conforme mencionado, a segunda etapa da pesquisa referiu-se ao estudo dos lexemas terminados em *-inho/-inha* no *corpus* de língua geral do NILC/São Carlos (Núcleo Interinstitucional de Linguística Computacional). Relacionado ao projeto AC/DC e

disponibilizado gratuitamente na Linguateca¹⁵, esse *corpus* se constitui de textos de língua portuguesa da variante brasileira, dos gêneros jornalístico, didático, epistolar e de redações de alunos.

Da busca no próprio site, extraíram-se duas listas com os 1.000 lexemas mais frequentes terminados em -inho e em -inha, somando-se 2.000 entradas. Manualmente, foram excluídos os nomes próprios e os casos de duplicação, o que totalizou 1.295 entradas, sendo: 604 com a terminação -inho, 601 com a terminação -inha e 90 com as duas terminações (*pobrezinho/pobrezinha*, *saidinho/saidinha*; *vizinho/vizinha*). Cada entrada lexical foi analisada, segundo as propriedades apresentadas no Quadro 2.

Quadro 2 – Propriedades dos lexemas terminados em -inho/-inha extraídos do *corpus*

Propriedades do lexema						Tipo de diminutivo				Exemplos
Lexema -inho/-inha	FOrg	NComp	FOrig	Dim	FDim	Tamanho	Intensidade	Juventude	Juízo de valor	Frase do <i>corpus</i>
cachorrinho	cachorro	-	-	+	+	+	-	-	+	Seu cachorrinho suja as calçadas. Só nado cachorrinho .
folhinha	folha	-	-	+	+	+	-	-	-	A cebolinha é uma folhinha verde usada para temperar alimentos. A festa não começava no domingo marcado pela folhinha .
quadrinho	quadro	-	-	+	+	+	-	-	+	Fui comprar um quadrinho de Picasso por US\$ 40 mil. No quadrinho seguinte, o gibi joga a culpa no governo.
tourinho	touro	-	-	+	-	-	-	+	+	O filme com a história de um tourinho e seu amigo.

Fonte: Elaboração própria

Assim como feito na análise dos dados dos dicionários, foram ignorados, para fins de descrição, 115 lexemas por serem nomes compostos (*abelha-rainha*, *tatu-galinha*) e/ou nomes com pseudossufixos, isto é, nomes que não apresentam o processo de derivação com os sufixos estudados (*ancinho*, *marinho*). Sendo assim, os 1.180 lexemas restantes foram analisados, considerando seus usos nos fragmentos extraídos do *corpus* e classificados da seguinte maneira: 1.116 atuam como diminutivos reais, e 107 como falsos diminutivos. Os usos de diminutivos reais receberam, em sua grande maioria (550/1.116), o rótulo de diminutivo por *tamanho*, seguido dos casos de diminutivo por *juízo de valor* (429/1.116). Ainda no Quadro 2, os seguintes lexemas obtiveram polo positivo para *juízo de valor*: *cachorrinho* (designação de afetividade com teor de ironia), *quadrinho* (designação de depreciação com teor de ironia) e *tourinho* (designação de afetividade). Apesar de não ter sido o objetivo deste trabalho, a análise dos lexemas em *corpus* – contrapondo ao seu estudo em dicionários – parece ser um ambiente bastante propício para pesquisas futuras, pois possibilita a avaliação mais precisa do uso do diminutivo para demarcar ironia, sarcasmo, como aponta o trabalho de Wick-Pedro (2018), e juízos de valor de maneira geral.

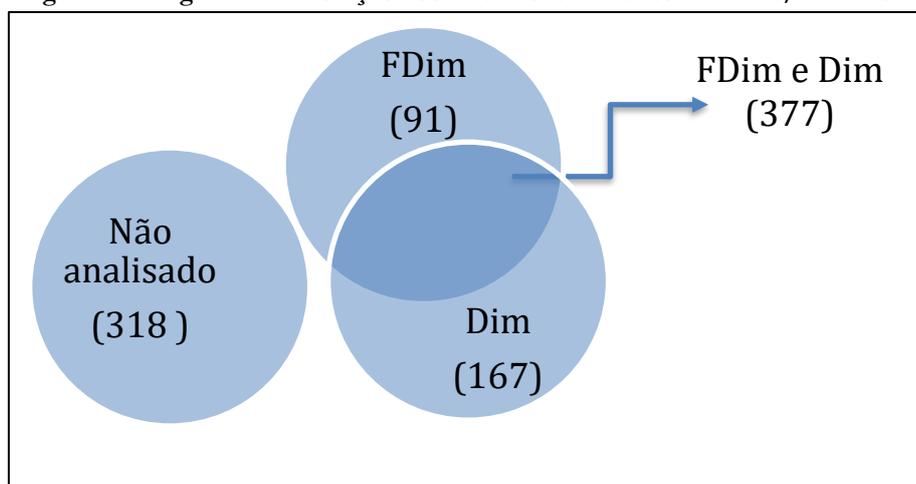
Para além da descrição e classificação dos dados em si, esperava-se, com a análise do *corpus*, encontrar novos casos de falsos diminutivos. Ao todo, houve a inclusão de apenas 11 lexemas que

¹⁵ A escolha pelo *Corpus* do NILC/São Carlos deveu-se à sua facilidade de acesso na época da primeira extração dos dados, em 2011.

não estavam nos dicionários (*biribinha, cortininha, bobinho, borrachinha, cabritinho, conchinha, cruzadinha, pombinhos, paradinha, tijolinho e verdinha*). Além dessas, outras 4 palavras foram anotadas, mas não incluídas na lista final, por serem compreendidas, neste trabalho, como usos muito específicos que remetem a marcas: *bisnaguinha, caçulinha, coelbinha e danoninho*.

Da junção dos dados extraídos dos dicionários e dos 11 lexemas anotados como falsos diminutivos encontrados no *corpus* do NILC, criou-se uma lista com 953 lexemas terminados em -inho/-inha, distribuídos da seguinte maneira: 318 casos não analisados, por se referirem a nomes compostos e/ou pseudossufixos; 167 lexemas anotados exclusivamente como diminutivos reais; 91 anotados exclusivamente como falsos diminutivos; e 377 lexemas que podem atuar como diminutivos reais e falsos diminutivos. A Figura 1 ilustra os dados levantados e analisados.

Figura 1 – Diagrama de anotação dos lexemas terminados em -inho/-inha do PB



Fonte: Elaboração própria

Como o objetivo desta pesquisa é o de avaliar o processamento de um dicionário eletrônico sobre os casos de falsos diminutivos do PB, a lista utilizada nesta investigação foi a de casos exclusivos de FDim (91) somada aos casos de Dim e FDim (377 lexemas), o que totalizou 468 lexemas.

Na próxima seção, serão apresentados os dados referentes ao processamento dessas palavras em *corpora* de língua geral e sua cobertura pelo dicionário DELAF-BR.

3 Tratamento dos falsos diminutivos: análise em um dicionário computacional

O léxico computacional, ou *dicionário*, é um elemento basilar para a grande maioria das ferramentas e aplicações de PLN. Segundo Muniz (2004, p. 5), trata-se de um recurso que contém os itens lexicais isolados (*lua, mel*) ou compostos (*lua de mel*) de uma língua e apresenta informações relacionadas a esses itens, tais como categorias gramaticais (*part-of-speech*) e valores morfossintáticos (*gênero, número, grau, pessoa, tempo, modo, etc.*).

Para discutir a respeito de como os dicionários computacionais tratam os casos de falsos diminutivos do PB, selecionou-se o formato de dicionário de formas flexionadas do DELA, denominado DELAF, que, como apresentam Finatto *et al.* (2019, p. 67), foi desenvolvido inicialmente para a língua francesa no Laboratoire d'Automatique Documentaire et Linguistique (LADL)¹⁶, e, posteriormente, elaborado para outras línguas naturais, como o inglês, o espanhol, o

¹⁶ Mais informações sobre o LADL disponíveis em: <http://infolingu.univ-mlv.fr/LADL/Historique.html>. Acesso em setembro de 2022.

árabe, o português, entre outras. O DELA é um dicionário computacional que abarca tanto dicionários de lemas (DELAS – formas simples; DELAC – formas compostas), quanto o dicionário de formas flexionadas (DELAF), sendo este último o interesse desta investigação.

O DELAF é um dicionário incorporado ao Unitex¹⁷ (PAUMIER, 2021), um *software* de código aberto, multiplataforma, plurilíngue, baseado em léxicos e gramáticas, utilizado para o processamento de *corpora*. Segundo Calcia (2015, p. 800), o Unitex é um

software [que] utiliza recursos linguísticos na forma de dicionários e gramáticas eletrônicas de grande cobertura para várias línguas (Espanhol, Inglês, Francês, Português, etc.) que descrevem as palavras simples e compostas, associando-as a um lema e a uma série de códigos gramaticais, semânticos e flexionais. Esses dicionários podem ser aplicados aos textos para a localização de padrões morfológicos, lexicais e sintáticos, remoção de ambiguidades e até mesmo para a etiquetagem de palavras simples e compostas, possibilitando aos usuários a busca de ocorrências por categoria gramatical ou ainda pelo lema das entradas lexicais. (CALCIA, 2015, p. 800).

Dessa forma, a fim de observar a cobertura dos casos de falsos diminutivos no dicionário computacional DELAF, foi utilizado o *software* Unitex para a obtenção dos seguintes dados: (i) análise da lista dos 468 lexemas anotados como falsos diminutivos por frequência em dois *corpora* de língua geral: o Folha94-02¹⁸ com textos do jornal Folha de S. Paulo entre os anos de 1994 e 2002, totalizando 135 milhões de ocorrências; e o Folha Kaggle¹⁹, com textos do mesmo jornal abrangendo o período de 2015 a 2017, com 167.053 notícias e 73 milhões de ocorrências. Essa etapa foi necessária para se ter noção dos casos de falsos diminutivos que realmente são recorrentes em língua portuguesa (já que aparentemente houve casos de falsos diminutivos relacionados a campos semânticos muito específicos – cf. Quadro 1); e (ii) análise da cobertura do DELAF-BR a partir da lista criada de 468 lexemas anotados como falsos diminutivos. É importante salientar que esses *corpora* foram selecionados por possuírem características comuns (ambos serem de fácil acesso para os pesquisadores e se constituem de textos do jornal Folha de S. Paulo) e por abarcarem recortes temporais distintos, o que possibilitaria uma análise comparativa de falsos diminutivos nesse gênero textual específico.

No que se refere à *primeira etapa*, inseriu-se no Unitex a lista de *falsos diminutivos* e foram processados os dois *corpora* selecionados para a obtenção dos dados relacionados à frequência de uso desses lexemas. A Tabela 2 apresenta os lexemas anotados como falsos diminutivos com frequência >70 nos *corpora*. Nessa lista, constam os lexemas que apresentam ao menos 1 (um) uso como falso diminutivo. Assim, casos de nomes próprios (*Bentinbo*, *Chacrinha*, *Felipinho*, *Mineirinho*), de verbos (*mantinha*) e de diminutivos exclusivamente reais (*bonitinha*, *casinha*, *pezinho*) foram retirados da análise.

Tabela 2 – Falsos diminutivos com frequência >70 em *corpora* do PB

FOLHA94-02 (313 formas)		FOLHA KAGGLE (284 formas)	
Lexema	Frequência	Lexema	Frequência
camisinha	2049	carrinho	556
cestinha	978	coxinha	371
cursinho	749	cursinho	342

¹⁷ Unitex disponível em: <http://www-igm.univ-mlv.fr/~unitex/>. Acesso em setembro de 2022.

¹⁸ *Corpus* obtido pela compilação dos CD-ROM vendidos pela Folha de S. Paulo com a maioria dos textos publicados naquele jornal entre os anos de 1994 e 2002.

¹⁹ *Corpus* Folha de S. Paulo disponível em: <https://www.kaggle.com/datasets/marlesson/news-of-the-site-folhaul>. Acesso em setembro de 2022.

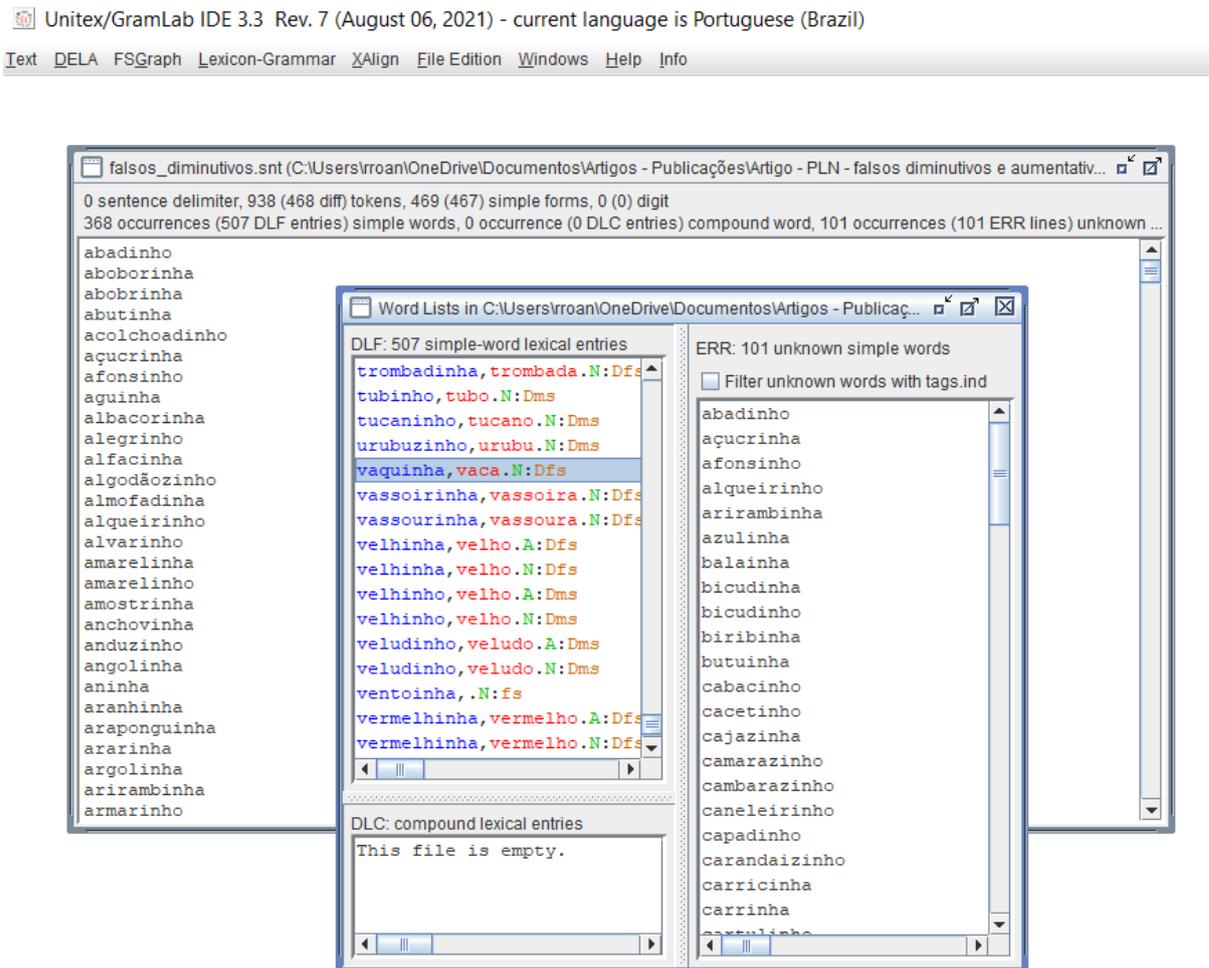
carrinho	692	jeitinho	292
caixinha	534	vaquinha	234
colarinho	477	caixinha	228
escolinha	451	cestinha	215
mauricinho	442	dobradinha	209
calcinha	425	camisinha	207
cafezinho	380	caipirinha	176
baixinho	379	colarinho	164
jeitinho	365	cafezinho	160
dobradinha	339	pegadinha	153
caipirinha	293	calcinha	153
mocinha	270	escolinha	152
bandeirinha	257	marchinha	135
safrinha	216	mocinha	126
mocinho	215	abobrinha	123
cebolinha	198	cebolinha	113
quadrinho	168	quadrinho	92
picadinho	155	safrinha	91
abobrinha	143	picadinho	90
salsinha	135	mocinho	88
amarelinho	132	puxadinho	78
escadinha	129	bandeirinha	75
salgadinho	129	casquinha	75
lanterninha	116	beijinho	74
vaquinha	113		
santinho	111		
patricinha	100		
cachorrinho	96		
tabelinha	96		
marchinha	93		
peixinho	80		
amarelinha	73		
beijinho	73		
canarinho	70		

Fonte: Elaboração própria

Os dados levantados são interessantes por demonstrarem mudanças significativas na frequência dos lexemas. A palavra *camisinha*, por exemplo, foi a mais frequente no *corpus* Folha94-02, mas aparece apenas na 9ª (nona) posição no *corpus* da Folha Kaggle, atrás de itens como *cestinha*, *vaquinha* e *coxinha*. Isso tem a ver, obviamente, com o tamanho de cada *corpus* (Folha94-02, 135 milhões de palavras; Folha Kaggle, 73 milhões de palavras) e com o período de criação e publicação dos textos (Folha94-02, de 1994 a 2002; Folha Kaggle, de 2015 a 2017).

Diante da lista geral de 468 falsos diminutivos, e levando-se em consideração os casos mais frequentes levantados em *corpora* (37 lexemas), realizou-se a *segunda etapa* de análise. Nela, verificou-se o processamento desses dados pelo dicionário computacional DELAF-BR, por meio do *software* Unitex (PAUMIER, 2021). A Figura 2 ilustra o processamento da lista e as informações sobre cada lexema segundo as informações do DELAF-BR.

Figura 2 – Processamento dos dados no Dicionário DELAF-BR no *software* Unitex²⁰



Fonte: Elaboração própria

Como se pode constatar na Figura 2, o *software* Unitex possibilita o processamento da lista de falsos diminutivos, gerando as informações sobre a classe e os valores morfossintáticos de cada lexema, segundo suas propriedades previstas no dicionário DELAF-BR. O lexema *vaquinha*, por exemplo, apresenta apenas uma entrada no dicionário, com as propriedades: *Nome* (substantivo), *Dfs*: diminutivo, feminino, singular. Embora contemple o uso diminutivo de *vaca* > *vaquinha*, o dicionário não prevê a sua entrada como falso diminutivo, que se refere à “coleta de dinheiro por um grupo de pessoas para o pagamento de uma despesa comum” (HOUAISS, 2009). Em vista disso, vê-se a necessidade de inclusão desse e de outros lexemas em seu uso não composicional. É importante salientar que o lexema é entendido como falso diminutivo no DELAF-BR quando apresenta uma entrada independente, sem a menção à sua forma diminutiva, tal como ocorre com *dobradinha*, que apresenta 3 entradas: (i) *dobradinha.N:fs*; (ii) *dobradinha,dobrado.A:Dfs*; e (iii) *dobradinha,dobrado.N:Dfs*. Em (i) tem-se o caso de falso diminutivo, por apresentar a informação *nome, feminino, singular* independente de sua relação diminutiva. Já em (ii) e (iii) *dobradinha* aparece como diminutivo de *dobrado* atuando como um adjetivo (A) e um substantivo (N), respectivamente.

²⁰ Como se verifica na Figura 2, cada entrada lexical apresenta: forma terminada em -inho/-inha, forma da palavra da qual deriva, classe de palavra (Nome, Adjetivo, Verbo), seguida das propriedades morfossintáticas, tais como: *D*, diminutivo; *f*, gênero feminino; *m*, gênero masculino; *s*, forma singular; *p*, forma plural; *P*, tempo verbal Presente, entre outras.

Considerando o exposto, ao processar-se a lista de 468 falsos diminutivos, obtiveram-se os seguintes resultados:

- 101 lexemas não estão descritos no dicionário eletrônico, dispostos na lista *unknown simple words* (*fezinhba, paradinba, pedalinbo, raspadinba, vestibulinbo*);
- 18 lexemas estão no dicionário com a informação morfossintática de sua atuação como verbo (*gatinba, gatinbar; mantinba, manter*);
- 318 lexemas estão no dicionário apenas relacionados ao uso como diminutivo real (*chuveirinbo, cruzadinba, pegadinba*);
- 31 lexemas estão contemplados no dicionário eletrônico como falsos diminutivos, sendo 22 casos anotados exclusivamente como falsos diminutivos (*colarinbo, enroladinbo, ribeirinbo, sombrinba*), 7, como falsos diminutivos e diminutivos reais (*bandeirinha, golfinbo, pracinba, quentinha*) e 2 como falsos diminutivos e atuação como verbo (*bacorinha e pintainbo*).²¹

Note-se, portanto, que o dicionário apresenta apenas 31 lexemas anotados como falsos diminutivos, ou seja, apenas 6,6% dos casos estão contemplados no dicionário computacional. Além disso, por ser reconhecido que muitos dos nomes admitidos na lista de falsos diminutivos são pouco frequentes em textos não especializados, tais como nomes de espécies de animais e plantas, foi analisada a lista de lexemas com frequência >70 (cf. Tabela 2) a partir dos dados retirados de *corpora* de textos jornalísticos do PB. Desses casos, dos 37 lexemas levantados, apenas 4 (10,8%) constam no DELAF-BR como entradas independentes (*bandeirinha, colarinbo, dobradinba e mauricinbo*). Todos os demais casos estão contemplados como formas diminutivas dos nomes dos quais derivam (*santinbo, santo.N:Dms*) e um caso (a palavra *patricinba*) é tido como uma forma desconhecida no dicionário.

A ausência da cobertura do fenômeno dos falsos diminutivos do PB em um dicionário eletrônico relevante no cenário do PLN no país corrobora a necessidade de construção de recursos linguísticos descritivos refinados para a ampliação desse e de outros repositórios léxicos que servem de base para distintas aplicações computacionais.

Considerações finais

Neste trabalho, são revisadas, aprofundadas e ampliadas listas de lexemas terminados em -inho/-inha do PB, além de ter sido investigado o processamento dos casos de falsos diminutivos por um dicionário computacional de considerável importância e utilização na área de PLN no Brasil.

Considerando os objetivos traçados e os dados descritos, pode-se afirmar que os lexemas terminados em -inho/-inha podem atuar como: (i) diminutivos reais, também nomeados *composicionais*, designando grau de *tamanho*, de *intensidade*, *juventude* e/ou *juízo de valor*, sendo essa última classe bastante relevante, principalmente na análise do uso desses lexemas; e (ii) falsos diminutivos, também conhecidos como *não composicionais*, relacionando-se aos mais variados campos semânticos, sobretudo: *botânica, ornitologia, ictiologia, sinonímia de aguardente e termo do esporte*.

O dicionário computacional analisado, DELAF-BR, demonstrou não possuir os dados relativos aos falsos diminutivos, processando apenas 31 dos 468 lexemas de maneira adequada, ou seja, apenas 6,6% dos casos. Considerando-se a lista de falsos diminutivos mais frequentes em

²¹ O DELAF-BR duplica ou multiplica as entradas dos lexemas a depender da sua atuação morfossintática, como exemplificado com o lexema *dobradinha*, que apresentada três entradas no dicionário. É devido aos casos de duplicação ou multiplicação de entradas que os valores apresentados no Unitex (cf. Figura 2) ultrapassam as 468 palavras da lista de falsos diminutivos.

corpora de textos jornalísticos do PB, o resultado não foi muito diferente: dos 37 falsos diminutivos anotados, o DELAF-BR cobriu apenas 4 casos (10,8%). Isso significa que o processo de seleção, análise e descrição desses lexemas é uma tarefa importante e necessária para a área dos estudos descritivos e para a criação de recursos léxicos que podem e devem ser utilizados na área de PLN.

A fim de aprofundar e ampliar o presente estudo, verifica-se a possibilidade do estabelecimento de uma agenda descritiva, com as seguintes propostas: revisão, ampliação e descrição da lista de aumentativos e falsos aumentativos do PB (*orelhão, pontilhão*), considerando, obviamente, suas especificidades; e estudo do uso de formas diminutivas (composicionais e não composicionais) em palavras e construções/expressões compostas, tal como ocorre em expressões cristalizadas: *tirar o cavalinho da chuva; dar uma mãozinha; pôr as manguinhas de fora*.

Portanto, pode-se afirmar que esta pesquisa contribui para o estado da arte, no que se refere às discussões teóricas a respeito dos lexemas terminados em *-inho/-inha* e, sobretudo, para a área de PLN, com a geração de uma lista de 468 falsos diminutivos, disponível para ser incluída aos dados do dicionário eletrônico analisado (DELAF-BR) e utilizada em outros projetos linguístico-computacionais.²²

Agradecimento

Parte deste trabalho foi realizado no âmbito do Centro de Inteligência Artificial da USP (C4AI - <http://c4ai.inova.usp.br/>), que tem o apoio da IBM e da FAPESP (processo 2019/07665-4). Este projeto também foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, que tem recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

Referências Bibliográficas

ALVES, Elisabeth. O diminutivo no português do Brasil: funcionalidade e tipologia. **Estudos Linguísticos** (São Paulo), v. XXXV, 2006.

ARMELIN, Paula Roberta. The non-compositional domain: diminutives and augmentatives in Brazilian Portuguese. **Estudos Linguísticos**, São Paulo, 43 (1), p. 395-410, 2014.

AZEREDO, José Carlos de. **Gramática Houaiss da Língua Portuguesa**. 2ª ed. São Paulo: Publifolha, 2008.

BECHARA, Evanildo. **Moderna gramática portuguesa**. 37ª ed. Rio de Janeiro: Nova Fronteira, 2009.

CALCIA, Nathalia Perussi.; KUCINSKAS, Adriele Beatriz; MUNIZ, Marcelo; NUNES, Maria das Graças Volpe; VALE, Oto Araújo. Révision et adaptation des dictionnaires et graphes de flexion d'Unitex-PB à la nouvelle orthographe du portugais. **3rd. UNITEX/ GRAMLAB WORKSHOP**, Tours, 2014.

CALCIA, Nathalia Perussi. O Acordo Ortográfico aplicado aos grafos e aos dicionários do Unitex. **Estudos Linguísticos** (São Paulo. 1978), v. 44, p. 799-811, 2015.

²² A lista completa dos falsos diminutivos do PB está disponível para utilização em outras investigações acadêmicas sem fins lucrativos: <https://repositorio.ufscar.br/handle/ufscar/18087>. Acesso em maio de 2023.

CUNHA, Celso; CINTRA, Lindley. **Nova gramática do português contemporâneo**. 5.ed. Rio de Janeiro: Lexikon, 2008.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário Aurélio da Língua Portuguesa**. Rio de Janeiro: Positivo, 2009.

FERREIRA, Marcelo; LOPES, Marcos. **Para conhecer Linguística Computacional**. São Paulo: Contexto, 2021.

FINATTO, Maria José Bocorny; VALE, Oto Araújo; LAPORTE, Éric. Reconhecimento do vocabulário de jornais populares brasileiros por um dicionário computacional de acesso livre. **Alfa**, São Paulo, v.63, n.1, p.63-80, 2019.

GMHP, GRUPO DE MORFOLOGIA HISTÓRICA DO PORTUGUÊS. Em busca de um método de investigação para os fenômenos diacrônicos. *In*: VIARO, Mário Eduardo. (org.). **Morfologia Histórica**. São Paulo: Cortez, 2014.

GONÇALVES, Carlos Alexandre. Flexão e derivação: o grau. *In*: Vieira, Silvia Rodrigues; Brandão, Silvia Figueiredo. (org.) **Ensino de gramática: descrição e uso**. 2. ed. São Paulo: Contexto, 2019.

GONÇALVES, Carlos Alexandre. **Iniciação aos estudos morfológicos: flexão e derivação**. São Paulo: Contexto, 2020.

HOUAISS, Antônio. **Dicionário Houaiss da Língua Portuguesa**. São Paulo: Editora Objetiva, 2009.

LOPES, Lucelene; DURAN, Magali; FERNANDES, Paulo; PARDO, Thiago. PortiLexicon-UD: a Portuguese Lexical Resource according to Universal Dependencies Model. **Proceedings of the 13th Edition of the Language Resources and Evaluation Conference**, Marselha, 2022.

MENDES, Ronald Beline. Diminutivos como marcadores de sexo/gênero. **Revista Linguística**, vol. 8, n. 1, 2012.

MORAES, Cívia; OLIVEIRA, Solange Mendes. Diminutivos: uma Análise Pragmática do Programa Vídeo Show. **Eletras (UTP)**, v. 20, 2010.

MUNIZ, Marcelo Caetano Martins. **A construção de recursos linguístico-computacionais para o português do Brasil: o projeto Unitex-PB**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional –Universidade de São Paulo, São Carlos, 2004.

PAUMIER, Sébastien. **Unitex 3.3: user Manual Unitex**. Paris: Université Paris-Est Marne-la-Vallée, 2021.

PETTER, Margarida Maria Taddoni. Morfologia. *In*: FIORIN, José Luiz (org.). **Introdução à Linguística: II. Princípios de análise**. 5. ed. São Paulo, 2021.

RODRIGUES, Roana; VALE, Oto Araújo. Calça, calcinha, calção: Falsos diminutivos e falsos aumentativos no português do Brasil. *In: Anais do II Congresso Brasileiro de Morfologia*. Rio de Janeiro, 2013.

RODRIGUES, Roana; SOUZA, Jackson Wilke Cruz; SANTOS, Roney Lira de Sales. Descrição linguística e aprendizado de máquina. **Cadernos De Estudos Linguísticos**, v. 64, 2022.

SILVA, Marcos Alexandre Rose; DIAS, Ana Luiza; ANACLETO, Junia Coutinho. Processing common sense knowledge to develop contextualized computer applications. *In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer: Berlin, Heidelberg, 2010.

VALE, Oto Araújo; BAPTISTA, Jorge. Novo dicionário de formas flexionadas do UnitexPB: avaliação da flexão verbal. **Proceedings of Brazilian symposium in information and human language technology**, Natal, 2015.

WICK-PEDRO, Gabriela. **COMENTCORPUS: Identificação e pistas linguísticas para detecção de ironia no português do Brasil**. Dissertação (Mestrado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2018.

Submetido em 15/04/2023

Aceito em 16/05/2023