

Parâmetros para anotação de gênero gramatical em nomes do português brasileiro

**Lucas Lima Gregório^{1,2}, Grazielle Santos Martins^{1,2}, Eulália Santana Freitas^{1,4},
Túlio Sousa de Gois^{1,3}, Raquel Meister Ko. Freitag^{1,2}**

¹Laboratório Multusuário de Informática e Documentação Linguística
Universidade Federal de Sergipe – Brazil
Didática II – 49.107-230 – São Cristóvão – SE – Brazil

²Departamento de Letras Vernáculas – Universidade Federal de Sergipe – UFS

³Departamento de Computação – Universidade Federal de Sergipe – UFS

⁴Departamento de Letras Estrangeiras – Universidade Federal de Sergipe – UFS

{oxen, graziufs, eulaliafreitas, tuliosg, rkofreitag}@academico.ufs.br

Abstract. This study investigates the attribution of grammatical gender in Brazilian Portuguese, with a focus on the Sergipan variety, using the Falares Sergipanos database and the spaCy tool. After automatic labeling and expert review, a significant discrepancy was observed between human and automatic classifications, highlighting error patterns related to gender stereotyping and technological limitations. Eight parameters were proposed to classify nouns in terms of grammatical gender according to morphological and syntactic features. The results contribute to the description of spoken Brazilian Portuguese, support the development of more inclusive language technologies and promote Open Science and replicability practices.

Resumo. Este estudo investiga a atribuição de gênero gramatical no português brasileiro, com foco na variedade sergipana, utilizando o banco de dados Falares Sergipanos e a ferramenta spaCy. Após a etiquetagem automática e revisão por especialistas, observou-se discrepância significativa entre as classificações humanas e automáticas, evidenciando padrões de erro relacionados à estereotipia de gênero e limitações tecnológicas. Foram propostos oito parâmetros para classificar substantivos quanto ao gênero gramatical segundo traços morfológicos e sintáticos. Os resultados contribuem para a descrição do português brasileiro falado, apoiam o desenvolvimento de tecnologias linguísticas mais inclusivas e promovem práticas de Ciência Aberta e reproduzibilidade.

1. Introdução

O gênero gramatical é um traço obrigatório de classificação de palavras no português brasileiro. Todas as palavras de base nominal têm gênero, e diferentemente do que se pensa no senso comum, não existe uma explicação lógica e baseada em sexo para isso. Ferramentas de processamento de linguagem natural, como o spaCy [Honnibal et al. 2020], atribuem rótulo de gênero gramatical às palavras do português, através de modelos pré-treinados em dados anotados.

Embora exista uma certa regularidade predizível no gênero gramatical relacionado à morfologia, há situações que demandam um ajuste mais refinado, com mais parâmetros acerca dos atributos relacionados à expressão do gênero gramatical.

Este estudo tem por objetivo identificar os padrões de frequência de gênero gramatical no Banco de Dados Falares Sergipanos [Freitag 2013], amostra Deslocamentos 2020 [Siqueira and Freitag]¹. São propostos oito parâmetros para os atributos de gênero para o tratamento computacional de um banco de dados sociolinguístico com etiquetagem morfossintática. A comparação entre a etiquetagem humana e a provida pela máquina visa a identificação de contextos que demandam ajustes nas regras. Além do resultado descritivo, o estudo provê dados autênticos para tarefas de aprendizado supervisionado, com dados rotulados quanto ao gênero no português brasileiro falado da variedade sergipana.

2. Frequencia de gênero gramatical e estereotipia

De maneira geral, pode-se predizer o gênero gramatical no português por meio de sua terminação: se termina com -o é masculino, e se termina com -a é feminino. No entanto, nem sempre essa correspondência é direta. A frequência de recorrência de um nome na sociedade pode atribuir à informação gramatical a sua configuração social; é o que ocorre com o gênero nas profissões expressas por nomes comuns de dois gêneros.

Enquanto há uma associação entre -o e masculino e -a e feminino quase categórica em nomes biformes e nos nomes substantivos que se referem a pessoas, no conjunto de nomes comuns de dois gêneros, a vogal temática evoca uma associação predominante com um dos gêneros, por razões que extrapolam o sistema linguístico. A regularidade da ocorrência e a frequência com que essa associação ocorre configuram a prototipicidade de gênero e essa informação também está codificada na gramática. Isso fica evidente no conjunto de nomes comuns de dois gêneros relativos a profissões, cuja vogal temática é -a, como em motorista, babá, dentista, frentista. Quando pensamos em motorista, a associação é com o gênero masculino, enquanto quando pensamos em babá, a associação é com o gênero feminino [Pinheiro and Freitag 2020]. Uma explicação para isso decorre das experiências com os estereótipos compartilhados na comunidade associando profissões e gêneros. No caso de nomes comuns de dois gêneros, na falta de informação gramatical de gênero nos substantivos, as pessoas escolhem um gênero a partir de representações mentais construídas por estereótipos de gêneros socialmente construídos [Pinheiro and Freitag 2020, Freitag 2024].

Estereótipos são representações, uma “imagem mental hipersimplificada de uma determinada categoria (normalmente) de indivíduo, instituição ou acontecimento, com-

¹A educação superior brasileira passou por uma reconfiguração de perfil em função de políticas públicas de expansão e interiorização, promovendo a diversificação do público que convive e interage no mesmo espaço. A documentação linguística no espaço universitário permite o estudo das dinâmicas de variação e mudança que ocorrem em função desta reconfiguração de perfil. Esta é a premissa para a constituição da amostra Deslocamentos, voltada ao estudo da mobilidade de estudantes da Universidade Federal de Sergipe, que dá suporte à descrição de diferentes fenômenos da fala sergipana, como ilustram [Correa 2019, Ribeiro 2019, Siqueira 2020, Cardoso 2021, Novais 2021, Pinheiro 2021, Rodrigues 2021, Silva 2021, Souza 2022, Siqueira 2025], dentre outros. Os procedimentos adotados na constituição desta amostra configuram um protocolo que permite reproduzibilidade em outras realidades, como [Bacelar 2025] e a realização de estudos contrastivos [Siqueira et al. 2023]. Por conta do detalhamento da descrição linguística, a amostra tem potencial de contribuir para o desenvolvimento de ferramentas de processamento automático da linguagem natural.

partilhada em aspectos essenciais, por grande número de pessoas” [Tajfel 2010]. Em princípio, não há uma predisposição favorável ou desfavorável na estereotipia. Os efeitos da estereotipia do gênero social podem emergir na concordância de gênero dos nomes na língua. A frequência de um nome comum de dois gêneros associado a um gênero específico configura pistas de estereotipia, que podem ser observadas também na comparação de frequências de nomes e gênero em uma amostra linguística de fala, tal como na amostra Deslocamentos 2020 [Siqueira and Freitag], que compõe o banco Falares Sergipanos [Freitag 2013].

3. Primeira etapa: concordância na classificação de gênero

3.1. Método

A amostra do banco de dados Falares Sergipanos, Deslocamentos 2020, composta por 60 entrevistas sociolinguísticas com, aproximadamente 1h de duração cada, gravadas em áudio, transcritas e alinhadas, foram submetidas à etiquetagem morfossintática por meio da biblioteca spaCy².

Após a etiquetagem, foram computadas as frequências das palavras classificadas como substantivo (NOUN) com dois elementos à esquerda e à direita, a fim da identificação do gênero gramatical e da classificação quanto à tipologia, seguindo a abordagem apresentada em [de Barros Santos et al. 2023]. Das 60 transcrições oriundas da amostra Deslocamentos foi constituído um *corpus* contendo 6.396 palavras classificadas automaticamente quanto ao gênero gramatical, que foram posteriormente anotadas por humanos juízes especialistas.

A primeira etapa do estudo consiste em comparar a classificação do sistema automático com uma base de referência anotada por humanos, visando identificar a frequência e o tipo de erros e calcular a concordância estatística entre as duas classificações. Após a execução da rotina de preparação e limpeza dos dados, passou-se à análise da distribuição de gênero³.

3.2. Resultados

Foi realizada uma contagem da frequência de cada categoria de gênero, tanto para a classificação automática quanto para a humana.

A classificação automática está descrita na Tabela 1 e a classificação humana na Tabela 2.

Table 1: Frequência de Gênero - Classificação Automática (spaCy)

Gênero	Frequência
Masculino	3.389
Feminino	2.997
Não aplicável	10

²O modelo utilizado para a classificação foi o pt_core_news_lg. Disponível em: https://spacy.io/models/pt#pt_core_news_lg.

³O *corpus* constituído para análise e os códigos desenvolvidos estão disponíveis no GitHub: <https://github.com/lamid-ufs/genero-gramatical>

Table 2: Frequência de Gênero - Classificação Humana

Gênero	Frequência
Feminino	2.649
Masculino	2.485
Não aplicável	1.126
Masculino e Feminino	136

O spaCy [Honnibal et al. 2020] classificou quase todos os itens, enquanto a anotação humana identificou 1.126 casos em que a atribuição de gênero não era aplicável, por não serem da classe dos substantivos, indicando a necessidade de ajustes do modelo disponível para o português brasileiro. A classificação humana identificou ainda nomes substantivos que oscilam entre masculino e feminino, aspecto a ser discutido à frente.

Excluídos os erros de categoria gramatical, para entender os erros do sistema quanto ao gênero gramatical, foram filtradas as ocorrências onde a classificação automática divergiu da humana. Uma matriz de confusão foi gerada para detalhar a natureza dessas discordâncias. O sistema tendeu a classificar palavras femininas como masculinas (104 casos) com mais frequência do que o inverso (48 casos), ver tabela 3. Tal resultado está em consonância com a estereotipia do masculino genérico.

Table 3: Matriz de confusão das discordâncias

Juízes spaCy	spaCy		
	Feminino	Masculino	Masculino e Feminino
Feminino	0	48	26
Masculino	104	0	106

A concordância entre as classificações de gênero foi avaliada através do coeficiente Kappa de Cohen (κ), que mensura o grau de concordância entre dois anotadores, levando em consideração concordâncias ocorridas ao acaso [Cohen 1960]. O resultado, $\kappa = 0.631$, mostra que, apesar dos erros identificados, o classificador automático apresenta uma concordância moderada [Landis and Koch 1977]. Após a filtragem das classificações de gênero com a remoção dos erros de categoria, o coeficiente foi para $\kappa = 0.895$, apresentando uma forte concordância. A definição de parâmetros de atributos do gênero gramatical pode auxiliar no refinamento e ajuste da categorização.

4. Segunda etapa: atributos de gênero gramatical

4.1. Método

Após a correção, seguiu-se o sistema de categorização e atributos [Freitag et al. 2021]. A categorização dos substantivos foi **masculino**, **feminino** e **masculino e feminino** (ver tabela 4). Este último rótulo de atributo abarca dois contextos: 1) nomes comuns de dois gêneros desacompanhados de especificadores ou modificadores e, 2) palavras de gênero vacilante. Nesta etapa, as análises ocorreram apenas nos dados anotados pelos juízes especialistas.

Os atributos da categorização são: 1) animacidade (+/- animado), 2) marcação de gênero na raiz, 3) marcação de gênero por desinência, 4) marcação de gênero por

Table 4: Categorização dos Substantivos

Categorização	Descrição
masculino	Substantivos de gênero único masculino, substantivos flexionados ou derivados em sua forma masculina ou substantivos comuns de dois gêneros acompanhados de especificadores ou modificadores masculinos
feminino	Substantivos de gênero único feminino, substantivos flexionados ou derivados em sua forma feminina ou substantivos comuns de dois gêneros acompanhados de especificadores ou modificadores femininos
masculino e feminino	Substantivos comuns de dois gêneros sem especificadores ou modificadores e substantivos de gênero vacilante

concordância, 5) marcação de gênero por derivação, 6) heteronímia, 7) forma: biforme ou uniforme, 8) tipo: comum, nome próprio, estrangeirismo, deverbal ou derivação imprópria.

A animacidade é uma das informações morfológicamente relevantes contidas nos radicais dos nomes, e distingue aqueles que designam seres vivos daqueles que designam entidades inanimadas [Villalva 2008]. As classificações referentes à marcação de gênero na raiz, por desinência, por concordância ou por derivação são excludentes entre si, o que significa que marcar sim para uma significa marcar não para as outras. Os substantivos classificados positivamente em relação à marcação de gênero por concordância foram aqueles que apresentam uma só forma para os dois gêneros, os chamados de comuns de dois gêneros [Rocha Lima 1973].

Os nomes uniformes comuns de dois gêneros são aqueles que apresentam uma única forma para se referir tanto ao gênero masculino quanto ao feminino. O gênero, nesse caso, é definido por pistas gramaticais, como adjetivos biforme e/ou artigos. Quando não foi possível identificar pistas de gênero, os substantivos foram classificados como sendo simultaneamente masculinos e femininos. Foram classificados dessa mesma forma os substantivos de gênero incerto ou vacilante, indiferentemente masculinos ou femininos [Moura Neves 2000], desprovidos de especificadores ou modificadores no contexto em que foram usados da entrevista.

Os substantivos classificados positivamente em relação à marcação de gênero por derivação foram aqueles em que a diferenciação entre os pares masculino e feminino dava-se por meio de sufixos derivacionais [Câmara Júnior 1992].

Alguns substantivos de gênero único com traço [+ animado] apresentam uma contraparte feminina por meio de palavra diferente [Bechara 2018]. Esses foram os substantivos classificados positivamente em relação à existência de heterônimos. O Minidicionário Contemporâneo da Língua Portuguesa [Caldas Aulete 2004] p. 419) define heteronímia como o “processo de formação de masculino ou feminino com palavra de radical diferente (p. ex., boi e vaca, cavalo e égua).” Para [Câmara Júnior 1992], p. 134), essa definição é incoerente, uma vez que “na descrição da flexão de gênero em português não há lugar para os chamados ‘nomes que variam em gênero por heteronímia.’” O autor ainda observa que, nas gramáticas, é comum encontrar a afirmação de que o substantivo *mul-*

her seria o feminino de *homem*. Entretanto, o autor considera essa definição equivocada, pois mulher é sempre um substantivo de gênero feminino, embora esteja semanticamente relacionada ao substantivo homem, que é privativamente masculino.

Table 5: Atributos dos Substantivos

Atributo	Descrição	Exemplo
animacidade	Distingue substantivos que designam seres vivos daqueles que designam entidades inanimadas [Vilalva 2007]	[+ animado] [+ humano]: <i>eleitores, lavrador</i> [+ animado] [- humano]: <i>camelo, caranguejo</i> [- animado]: <i>igrejas, pedra, norte</i>
marcação de gênero na raiz	Substantivos de gênero único [Lima 2022]	[+ animado] [+ humano] <i>membros, indivíduo</i> [+ animado] [- humano] <i>camelo, caranguejo</i> [- animado] [- humano]: <i>percurso, economia, incidentes</i>
marcação de gênero por desinênciā	Aqueles que para formar o feminino recebem o sufixo flexional -a com supressão da vogal temática, e seus alomorfes [Câmara Júnior 1992]	biformes: <i>amigo-amiga</i>
marcação de gênero por concordância	Apresentam uma só forma para os dois gêneros [Lima 2022], as marcas de concordância de elemento adjacente diferenciam	<i>agentes, turista e atleta</i>
marcação de gênero por derivação	Diferenciação de gênero por sufixo derivacional [Câmara Júnior 1992]	<i>herói, galinha e atriz</i>
heteronímia	Apresentam contraparte feminina por meio de palavra diferente [Bechara 2018]	<i>painho</i> (heterônimo <i>mainha</i>), <i>freiras</i> (heterônimo <i>freis</i>), <i>bode</i> (heterônimo <i>cabra</i>), <i>padre</i> (heterônimo <i>madre</i>).
forma	biformes: apresentam duas formas, uma para o masculino e outra para o feminino; uniformes: apresentam apenas uma forma	biforme: <i>menino-menina, aluno-aluna, professor-professora, juiz (juíza)</i> uniforme: <i>lugar, cadeira, livro</i>
tipo	Subclassificações de base semântica ou morfológica: comum, nome próprio, estrangeirismo, deverbal, derivação imprópria	comuns: <i>cidade, distopia, ouro</i> próprios: <i>Vitória, Yuri e Glória</i> estrangeirismos: <i>shows, iceberg e handball</i> deverbais: <i>impacto, reprodução, locomoção, globalização</i> impróprios: <i>o útil ao agradável</i>

4.2. Resultados

A contagem das ocorrências quanto aos parâmetros de atributos controlados, em relação à animacidade, aponta para o predomínio do gênero feminino em palavras com o traço [- animado], 46,5% contra 38,3% masculino e 0,5% masc./fem., enquanto palavras com o traço [+ animado] apresentam predomínio de palavras de gênero masculino, 8,8% contra 3,8% feminino e 2% masc./femin. Como substantivos constituem o maior contingente de palavras, em valores absolutos, pode-se dizer que o gênero grammatical predominante nesta amostra é o feminino, em uma tendência já identificada em outros estudos [Schwindt 2020]. Também observa-se que os substantivos [+ animado] são os que apresentam percentual maior de gênero masculino e feminino, sugerindo que o efeito da variância de gênero está associado a este atributo.

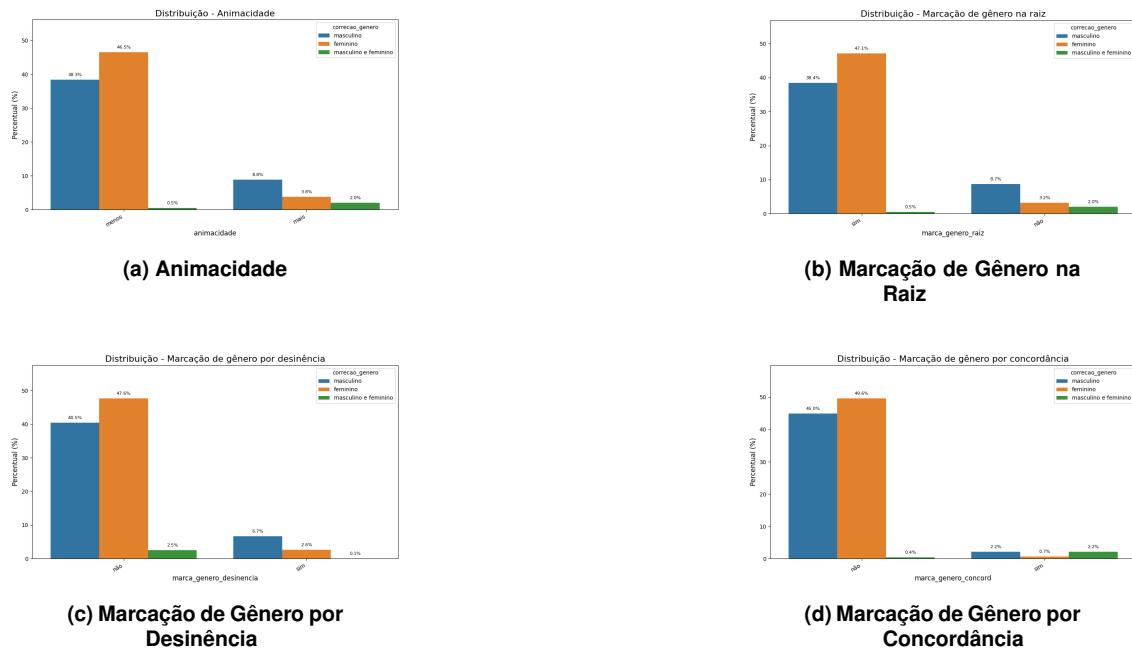


Figure 1: Distribuição dos atributos em relação à animacidade e parâmetros formais

A observação do tipo aponta que não há diferença entre masculino 36,5% e feminino 36,4% em substantivos comuns; no entanto, na formação de substantivos dever-bais, o predomínio feminino é expressivo, com 12,8%, contra 7,2% do masculino. Já as palavras de origem estrangeiras tendem a ser marcadas no masculino 1,8%, contra 0,3% do feminino. Este resultado reflete tendências identificadas em estudos anteriores [Surreaux and Schwindt 2021], sugerindo a especialização de gênero em situações específicas de formação de palavras. Os nomes que apresentam marcação masculina e feminina predominam na classe dos comuns, sendo pouco frequentes nas demais classes.

Em relação à marcação de gênero, identificamos uma polarização entre marcação de gênero na raiz e feminino, 47,1% contra 38,4% do masculino, e não marcação de gênero na raiz e masculino, 8,7% contra 3,2% do feminino. Nomes masculinos e femininos também não apresentam a marcação de gênero na raiz. A mesma tendência se repete nos substantivos em que a marcação de gênero se dá por desinéncia, com 47,6% feminino contra 40,5% masculino. Nomes masculinos e femininos também não são marcados

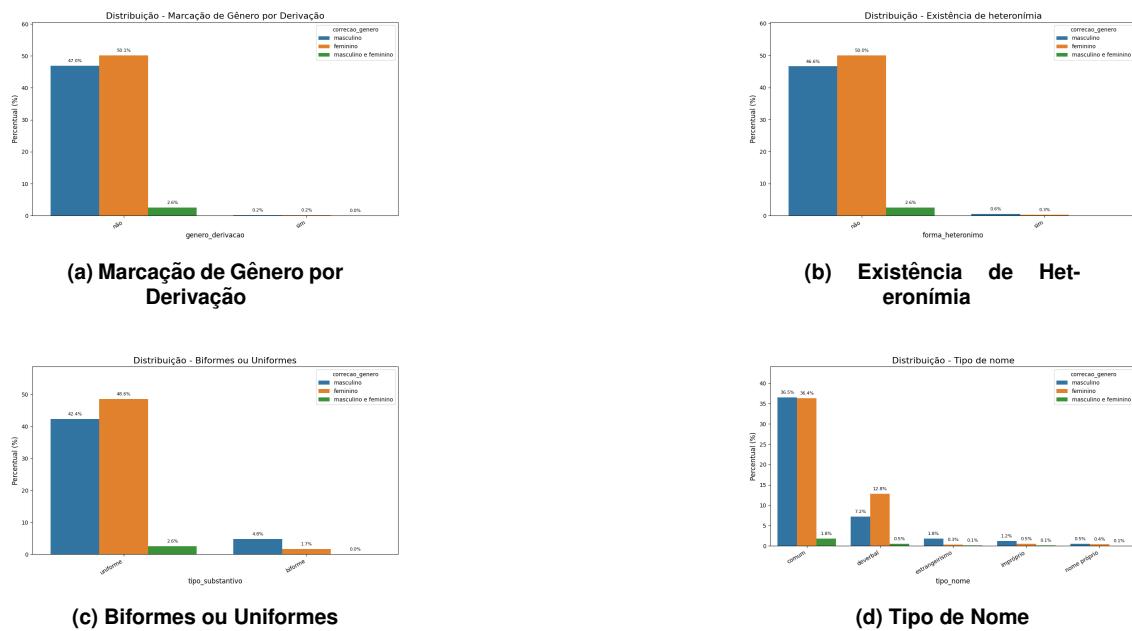


Figure 2: Análise da distribuição de gênero por tipo e formação do substantivo

na desinência. A marcação de gênero por derivação é pouco expressiva, e não apresenta diferença entre masculino e feminino, 0,2%. Já a marcação de gênero por processo de concordância é a estratégia predominante para diferenciar nomes masculinos e femininos (gênero vacilante, comuns de dois gêneros), enquanto a concordância no masculino genérico é predominante 6,6% contra 0,7%.

5. Conclusão

Os resultados da categorização do gênero grammatical contribuem para a ampliação da descrição do português brasileiro falado, especialmente da variedade sergipana.

A categorização dos substantivos, com base nos parâmetros de animacidade, marcação por raiz, por flexão, por derivação, por concordância, e os tipos morfológicos pode contribuir para o refinamento do processo de etiquetagem de gênero grammatical de ferramentas como o spaCy, que embora tenha um excelente desempenho em comparação à classificação humana, ainda apresenta espaço para uma concordância categórica, dado que gênero grammatical é obrigatório no português brasileiro, e essa marcação precisa ser considerada no desenvolvimento de aplicações em PLN. O predomínio do feminino em nomes [- animado] e a persistência do masculino genérico como “*default*” em classificações automáticas ilustram tensões entre norma grammatical, uso real e estereótipos sociais.

Estes resultados contribuem para o debate sobre linguagem inclusiva e comunicação não sexista. A identificação de padrões sistemáticos de erro na classificação automática, especialmente a tendência de rotular como masculinas palavras de uso ambíguo ou originalmente femininas, evidencia a necessidade de ajustes nos modelos de processamento de linguagem natural para que reflitam mais fielmente a diversidade do português falado, evitem reforçar estereótipos de gênero, e contribuam para o desenvolvimento de soluções tecnológicas mais justas e responsivas a demandas sociais por equidade de gênero.

Agradecimentos

Este estudo é vinculado aos projetos *Emergência de gênero não-binário: efeitos linguísticos, sociais e cognitivos*, financiado pelo edital Universal 2023 e pelo edital de Internacionalização 2023 do CNPq, e *MultiLingualGender: Multilingual perspective in Gender and Language. Realization and processing of gender in different Romance languages: an interdisciplinary approach for the field of education and public communication*, financiado pela chamada HORIZON-MSCA-2023-SE-01, da European Research Executive Agency.

References

- Bacelar, H. Q. (2025). A gente eu acho que tá correto também, mas não tão correto?: uso e avaliação de pronomes de primeira pessoa do plural por estudantes da ufsc. Dissertação (mestrado em linguística), Universidade Federal de Santa Catarina.
- Bechara, E. (2018). *Moderna gramática portuguesa*. Nova Fronteira, Rio de Janeiro, 39 edition.
- Caldas Aulete, F. J. (2004). *Minidicionário contemporâneo da língua portuguesa*. Nova Fronteira, Rio de Janeiro.
- Câmara Júnior, J. M. (1992). *Estrutura da língua portuguesa*. Vozes.
- Cardoso, P. B. (2021). Efeitos linguísticos e paralinguísticos na inferência dos sentidos indicados por (eu) acho que em entrevistas sociolinguísticas. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Correa, T. R. A. (2019). A variação na realização de /t/ e /d/ na comunidade de práticas da ufs: mobilidade e integração. Dissertação (mestrado em estudos linguísticos), Universidade Federal de Sergipe.
- de Barros Santos, V., Freitag, R. M. K., do Nascimento, R., Silva, E. S., and de Gois, T. S. (2023). Linguagem inclusiva e comunicação não sexista na universidade federal de sergipe. *Revista do GELNE*, 25(3):e32554–e32554.
- Freitag, R. (2024). *Não existe linguagem neutra!: gênero na sociedade e na gramática do português brasileiro*. Editora Contexto, São Paulo.
- Freitag, R., Tejada, J., Pinheiro, B., and Cardoso, P. (2021). Função na língua, generalização e reproduzibilidade. *Revista da ABRALIN*, pages 1–27.
- Freitag, R. M. K. (2013). Banco de dados falares sergipanos. *Working Papers em Linguística*, 14(2):156–164.
- Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Moura Neves, M. H. (2000). *Gramática de usos do português*. Editora da UNESP, São Paulo.

- Novais, V. S. (2021). Variação na concordância verbal de terceira pessoa do plural na fala de universitários sergipanos. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão, SE.
- Pinheiro, B. F. M. (2021). Pistas linguísticas e paralinguísticas para os sentidos diminutivos. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Pinheiro, B. F. M. and Freitag, R. M. K. (2020). Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência. *Linguística*, 16(1):85–107.
- Ribeiro, C. C. S. (2019). Deslocamento geográfico e padrões de uso linguístico: a variação entre as preposições em “ni” na comunidade de práticas da universidade federal de sergipe. Dissertação (mestrado em letras), Universidade Federal de Sergipe.
- Rocha Lima, C. H. d. (1973). *Gramática normativa da língua portuguesa*. J. Olympio.
- Rodrigues, F. G. C. (2021). Variação na regência de complementos locativos de verbos de movimento na fala de universitários da ufs. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão, SE.
- Schwindt, L. C. (2020). Predizibilidade da marcação de gênero em substantivos no português brasileiro. In Carvalho, D. and Brito, D., editors, *Gênero e língua(gem): formas e usos*, volume 1, pages 279–294. Editora da UFBA, Salvador / Bahia.
- Silva, L. S. (2021). Análise acústica ou de oitiva? contribuições para o estudo da palatização em sergipe. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Siqueira, J. (2020). Variação no preenchimento da posição determinante antes de possessivos pré-nominais: padrões dialetais e contatos. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Siqueira, J. (2025). *Covariação morfossintática no português brasileiro: identificação dialetal de estudantes da Universidade Federal de Sergipe*. Tese (doutorado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Siqueira, M. and Freitag, R. M. K. O controle da mobilidade na constituição de amostras sociolinguísticas. In *LABPEC (no prelo)*, pages 1–20.
- Siqueira, M., Sousa, M. D. A. F., and Rodrigues, F. G. C. (2023). Sistematizando padrões dialetais morfossintáticos: Mobilidade e contato. In Freitag, R. M. K. and Savedra, M. M. G., editors, *Mobilidades e Contatos Linguísticos no Brasil*, pages 165–188. Blucher, São Paulo.
- Souza, V. R. A. (2022). Monotongação dos ditongos decrescentes orais [o], [e], [a] e [o] na fala e na leitura em voz alta de universitários sergipanos. Dissertação (mestrado em letras), Universidade Federal de Sergipe, São Cristóvão.
- Surreaux, P. and Schwindt, L. C. (2021). Marcação de gênero gramatical em formações novas em português brasileiro. *Estudos Linguísticos e Literários*, (72):390–414.
- Tajfel, H. (2010). *Social identity and intergroup relations*, volume 7. Cambridge University Press.