

Eu, Tu, Ele, Ela, Elu, Nós, Vós, Eles, Elas, Elus por um Modelo de Linguagem Neutra

Washington Roberto Lopes¹, Be Zilberman¹, Bruna Magrini da Cruz¹,
Leticia dos Santos Monte Cruz¹, Rafaella Alves Lucena Gomes¹,
Renata Wassermann¹, Sarajane Marques Peres¹, Valdinei Freire¹

¹Comitê de Inclusão e Diversidade – Centro de Inteligência Artificial (C4AI)
Universidade de São Paulo
05508-020 - São Paulo - SP - Brasil

{renata.wassermann,sarajane,valdinei.freire}@usp.br

Abstract. *The neutral language is at the center of discussions surrounding inclusion and the fight against gender bias. Based on gender neutralization, it can involve the addition of new gender-neutral elements to a language or the prioritization of writing in a neutral syntax. Both approaches are automatically processable and can be incorporated within the scope of natural language processing. This article presents an initiative to optimize a language model focused on translating sentences from traditional Portuguese into neutral language, considering the new gender-neutral elements. For this, a bilingual corpus was constructed, encompassing manually translated paragraphs from news articles, words and phrases from an official guide on neutral language, as well as automatically generated sentences. The results obtained with the optimized language models demonstrate the feasibility of generating inclusive language models.*

Resumo. *A linguagem neutra está no centro de discussões sobre inclusão e combate a vieses de gênero. Pautada na neutralização de gênero, ela é caracterizada pela adição de novos elementos de gênero neutro em uma língua, ou pela priorização da escrita em sintaxe neutra. Ambas as formas são processáveis automaticamente e podem ser tratadas no escopo do processamento de linguagem natural. Este artigo apresenta uma iniciativa para otimizar um modelo de linguagem para traduzir sentenças do português oficial para a linguagem neutra, considerando os novos elementos de gênero neutro. Um corpus bilíngue foi criado para o treinamento, contendo traduções manuais de parágrafos de notícias, palavras e frases de um guia oficial de linguagem neutra, além de sentenças geradas automaticamente. Os resultados com os modelos otimizados mostram que a geração de modelos de linguagem inclusivos são factíveis.*

1. Introdução

O alcance de uma sociedade justa e igualitária depende da conscientização das pessoas sobre os problemas enfrentados por determinados grupos sociais. Conscientização está ligada à necessidade de educação, conhecimento e aceitação de responsabilidades. Quando um problema não é discutido, para a maioria que não sofre as consequências daquele problema, ele não existe. Discutir pressupõe falar, falar pressupõe usar uma língua. A língua é vetor de mudanças e tem uma importância sem igual para a sociedade.

Atualmente, a sociedade internacional tem testemunhado avanços tecnológicos advindos da área de processamento de linguagem natural, impulsionada pelos grandes modelos de linguagem [Liu et al. 2023, Chang et al. 2023, Carmo et al. 2020]. Construídos a partir do aprendizado indutivo, implementado via aprendizado de máquina, tais modelos têm se mostrado eficientes para diversas tarefas. Esses modelos são vetores que proporcionam a escalabilidade de muitas ações. O poder desses modelos de linguagem aliados ao poder transformador de uma língua criam um arcabouço ferramental de enorme potencial para a promoção da justiça e da igualdade.

A inserção da linguagem neutra no contexto dos modelos de linguagem torna-se não só importante como essencial [Cassiano 2023]. A linguagem neutra é estabelecida como um processo de reestruturação e questionamento de uma língua que, sob alguns ângulos, admite vieses. Com base na neutralização de gênero, a linguagem neutra é discutida como ferramenta social em diferentes países onde a língua oficial faz distinção de gênero. A linguagem neutra pode ser estabelecida sob duas vertentes. A primeira, também conhecida como linguagem inclusiva, estabelece uma prática de priorização da escrita em sintaxe neutra. A segunda, mais específica e também mais polêmica na sociedade, é pautada na adição de novos elementos de gênero neutro em uma língua.

Este artigo apresenta um estudo sobre a otimização de um modelo de linguagem para traduzir sentenças do português oficial para a linguagem neutra, considerando a inserção de novos elementos linguísticos para a língua portuguesa. Como exemplo do objetivo deste estudo, considere os casos de tradução mostrados na Figura 1. Nesta figura, o quadro do lado esquerdo especifica a entrada para o tradutor em português oficial. O quadro do lado direito mostra a tradução desejada, usando linguagem neutra.

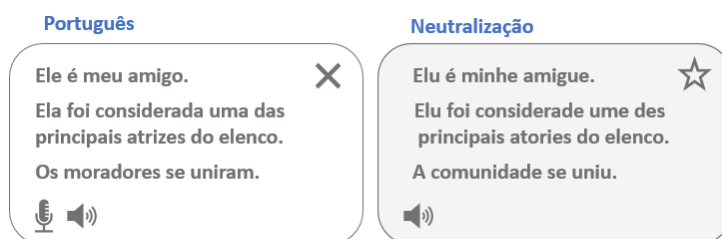


Figura 1. Exemplo de tradução do português oficial para a linguagem neutra, usando novos elementos linguísticos para neutralização de gênero. As traduções das duas primeiras sentenças envolve a linguagem neutra com adição de novos elementos linguísticos. A tradução da terceira sentença usa a sintaxe neutra. Figura inspirada na interface do serviço Google Tradutor.

Para desenvolvimento deste estudo, o modelo de linguagem pré-treinado PTT5 [Carmo et al. 2020] foi utilizado. Três instâncias de otimização do modelo foram obtidas a partir do uso de três corpora de treinamento: (a) um corpus baseado em notícias jornalísticas; (b) um corpus baseado em palavras e sentenças usadas em um guia oficial para a linguagem neutra; (c) um corpus baseado em sentenças sintéticas. A implementação da tradução do português oficial para a linguagem neutra traz desafios que, para além de uma justificativa de uso social, estabelece um desafio técnico para o desenvolvimento dos modelos, visto as dificuldades contextuais impostas pelo uso de novos elementos linguísticos e a falta de volume de dados textuais produzidos na linguagem neutra. O conteúdo deste artigo está organizado como segue: os contextos e conceitos sociais e técnicos pertinentes

ao estudo são apresentados na Seção 2; estudos relacionados são apresentados na Seção 3; a Seção 4 discute o problema de tradução abordado; o método usado no estudo é apresentado na Seção 5; os resultados e análises decorrentes estão organizados na Seção 6; finalmente, a Seção 7 conclui o artigo apresentando as considerações finais do estudo.

2. Contexto e conceitos

2.1. Linguagem Neutra - justificativa social

A linguagem neutra é um processo de reestruturação e questionamento da língua que pauta, por meio do uso de um gênero neutro, a inclusão de pessoas não-binárias, intersexo, de gênero não-conformante e indeterminado. Além disso, o uso da linguagem neutra questiona o que se conhece como “masculino genérico”, definido na linguística como o uso do gênero gramatical masculino para denotar o gênero humano (designando homens e mulheres) [Mäder 2015]. Tal conceito ainda remete ao que se conhece como “gênero não marcado” ou “extensivo” [Mäder 2015]. De forma minoritária, há línguas que aplicam o “feminino genérico”. No português, o “feminino genérico” ocorre culturalmente ao designar, por exemplo, profissões estereotipicamente femininas. A pauta construída sobre o desenvolvimento e disseminação da linguagem neutra não é nova, tendo surgido com os movimentos feministas, LGBTQIAPN+ e étnicos. Vieses de gênero associam conceitos a gêneros. Por exemplo, funções profissionais consideradas femininas ou masculinas, atividades mais atribuídas a mulheres ou homens e a exclusão de pessoas com identidades de gênero que não correspondem aos gêneros estabelecidos.

Para compreender o que é a linguagem neutra, se faz necessário discutir o conceito de gênero como uma construção social, que é aprendida e ensinada para os indivíduos, se consistindo em padrões de comportamento e aparência que são atribuídos a um gênero e ao outro. A linguagem neutra surge com o objetivo de readequar a utilização de pronomes de gênero na língua portuguesa e em diversas outras línguas. Ela ressignifica preconceitos de gênero, evitando os vieses existentes em várias línguas faladas no mundo, e também a exclusão de pessoas por meio da linguagem, com base em sua identidade de gênero.

Os benefícios da linguagem neutra não são unanimidade na sociedade. Há argumentos contra a sua adoção. Um dos pontos destacados é que as novas grafias tornam a comunicação mais complexa, de forma a dificultar seu aprendizado e compreensão por pessoas que não têm um bom conhecimento da língua. No Brasil, há discussões legais sobre a adoção da linguagem neutra. Em 2021, foi publicada uma portaria proibindo o uso da linguagem neutra em projetos financiados pela Lei Rouanet. A mesma portaria foi revogada posteriormente, mas sua existência ilustra uma discussão jurídica sobre o tema. Além disso, concursos públicos e vestibulares pressupõe o uso da norma culta do português, o que desmotiva o uso da linguagem neutra em redações, por exemplo.

2.2. Linguagem Neutra - aspectos teóricos

As bases teóricas da linguagem neutra usadas neste estudo são oriundas do Guia da Linguagem Neutra [Cassiano 2023]¹, idealizado por Ophelia Cassiano, pessoa não-binária. É fundamental reconhecer que a linguagem neutra não é oficializada e que existem diferentes abordagens e manuais disponíveis. O Guia da Linguagem Neutra de Ophelia Cassiano

¹O Guia da Linguagem Neutra está disponível no link <https://medium.com/guia-para-linguagem-neutra-pt-br> e está em constante desenvolvimento e atualização.

é amplamente aceito pela comunidade não-binária no Brasil e possui uma estrutura gramatical sólida e organizada. Ele é o maior e mais completo conteúdo que existe dentro da língua portuguesa no assunto, e é o mais acessado da América Latina e países lusófonos.

A compreensão de um estudo que incorpora a linguagem neutra em um modelo de linguagem requer um contato com os principais aspectos teóricos da linguagem. Desta forma, esta seção é dedicada a transpor para este texto uma pequena parte do conteúdo do Guia da Linguagem Neutra, selecionada para ilustrar o que foi incorporado nas ações desenvolvidas no presente estudo. O conteúdo trabalhado neste estudo é composto pelos aspectos decorrentes da inserção de elementos neutros na gramática da língua portuguesa, e apenas tangencia a readequação do discurso a partir da sintaxe neutra.

As adições (e substituições) abrangidas pelo primeiro aspecto da linguagem neutra são principalmente representados pelo pronome neutro ELU. O pronome ELU é a variação neutra dos pronomes “ela” e “ele”. Sua pronúncia foi estabelecida de forma flexível como “é-lu” ou “ê-lu”, e facilita a aquisição e uso da língua por pessoas de diferentes sotaques brasileiros. Apesar de existir uma grande variação de pronomes neutros (ile, ilu, el, ili, éle) e diferentes preferências entre as pessoas que os utilizam, ELU foi o escolhido no guia. A partir da adição deste pronome na língua, outros elementos se tornam possíveis e adequados: contrações (DELU, DESSU, NESSU, AE, DE, NE, PELE, DUME, PRE), pronomes demonstrativos (ESSU, ESTU, AQUELU), pronomes possessivos (MINHE, SUE, TUE, NOSSE); desinência para o final das palavras (-a/-o, -ca/-co, -ga/-go, -ça/-ço, -r/-ra, -ãos, -ães, -ões/-ãs, -oas, -onas; +E, +IES, +QUE, +GUE, +CE); pronome pessoal oblíquo (LE); numeral (UME e DUES), heterônimos (-mãe/-pai, +NÃE); termos biformes (-bom/-boa, +BOE); artigos definidos (Ê/ES). A forma pura do artigo é E (sem acento), mas o uso do circunflexo no singular é essencial para diferenciá-lo da conjunção “e”. O plural do artigo ES não tem acento, pois não se assemelha à conjunção. Sua forma pura sem acento também é usada dentro das contrações. Exemplos de uso desses elementos linguísticos são apresentados na Tabela 1.

Tabela 1. Exemplos da aplicação dos elementos neutros adicionados à língua portuguesa. Cores nas palavras indicam as adições e substituições requeridas. PT: português; LN.: linguagem neutra.

PT	O gato arranhou o braço dela .	Mandei mensagem para aquela gatinha .
LN	O gato arranhou o braço delu .	Mandei mensagem para aquele gatinhe .
PT	trabalhador (a)	trabalhadores (as)
LN	trabalhadore e	trabalhadorie s
PT	O Carlos e a Ana são namorados .	Conheci um mochileiro .
LN	Ê Carlos e ê Ana são namorades .	Conheci ume mochileire .
PT	m aternidade / p aternidade	pr íncipe / pr incesa
LN	n aternidade	pr íncise

A aplicação da sintaxe neutra é apresentada no Guia da Linguagem Neutra [Cassiano 2023] como um método de reorganização de uma sentença de modo a não incluir gênero feminino ou masculino. A prática da sintaxe neutra é mais facilmente aplicável à linguagem escrita formal, na elaboração de textos ou discursos. No guia, entretanto, é estimulado o uso da linguagem neutra em seu aspecto aditivo. A sintaxe neutra também é conhecida como linguagem inclusiva. A composição da sintaxe neutra

se dá na preferência pelo uso de: coletivos neutros (humanidade); substantivos uniformes de apenas um gênero (celebridade); substantivos uniformes comuns de dois gêneros (cliente); pronomes indefinidos (alguém); adjetivos sem gênero (alegre); eliminação de artigos e pronomes; uso da voz passiva e gerúndio; exclusão de definições sexistas (unissex). Exemplos desses elementos são apresentados na Tabela 2.

Tabela 2. Exemplos da sintaxe neutra [Cassiano 2023]. Cores nas palavras indicam as reformulações requeridas. PT: português; LN.: linguagem neutra.

PT	Reunião de país .	Que menino mais lindo.
LN	Reunião de família .	Que criança mais linda.
PT	Venha sozinho .	Não encontrei elas .
LN	Não traga acompanhantes .	Encontrei ninguém .
PT	Você é mentiroso .	Hormônio feminino / masculino .
LN	Você não é confiável .	Hormônio específico .
PT	A Ariel saiu de casa com a Cameron .	Estou preocupado com isso.
LN	Ariel saiu de casa com Cameron .	Isso está me preocupando .

2.3. Modelos de linguagem

Neste estudo foi utilizado um modelo de linguagem neural pré-treinado, baseado no modelo T5 (Text-To-Text-Transfer-Transformer) [Raffel et al. 2020], chamado de PTT5 [Carmo et al. 2020]. O modelo PTT5, assim como seu precursor T5, são modelos de linguagem baseados na arquitetura Transformer [Vaswani et al. 2017]. Transformer é uma arquitetura de rede neural composta por dois módulos, o *encoder*² e o *decoder*³, e tem como sua principal característica a aplicação de mecanismos de atenção que atribuem pesos a partes de uma sequência sob processamento na arquitetura, o que permite capturar dependências de longo alcance no processamento de uma sequência de símbolos. No caso de tarefas de processamento de linguagem natural, tal capacidade permite processar o contexto no qual as palavras estão inseridas, sendo importante para absorver o significado de uma palavra em uma sentença. A arquitetura Transformer é considerada o estado da arte para a resolução de tarefas de processamento de linguagem natural.

A ideia de proposição do modelo de linguagem T5 é tratar um problema de processamento de texto como um problema do tipo “texto-para-texto”. Desta forma, o modelo recebe como entrada um texto e produz um novo texto como saída [Raffel et al. 2020]. Considerando esta maneira de modelar um problema de processamento de linguagem natural, um mesmo modelo de linguagem pode ser usado para diferentes tarefas, a partir de um refinamento de seu aprendizado orientado à tarefa alvo no momento. Essa estratégia é ideal, por exemplo, para o objetivo de tradução apresentado no presente estudo. Contudo, um modelo pode ser mais bem aproveitado se ele for especialmente otimizado para tratar uma língua em especial antes de ser aplicado a uma tarefa naquela língua. O modelo PTT5 é uma otimização do modelo T5 a partir do uso do corpus BrWac (Brazilian Web as Corpus) [Wagner Filho et al. 2018] em um processo de refinamento do aprendizado (*fine-tuning*). De acordo com os proponentes do modelo PTT5 [Carmo et al. 2020],

²Encoder: estrutura que mapeia uma sequência de representação de símbolos para uma sequência de representação contínua.

³Decoder: estrutura que gera uma sequência de símbolos, um elemento por vez.

o pré-treinamento monolíngue do modelo apresenta um desempenho superior em tarefas específicas de processamento da língua portuguesa quando comparado ao desempenho do modelo pré-treinado em corpora multilíngues. A Figura 2 contextualiza a aplicação do modelo PTT5 no problema de tradução do português oficial para a linguagem neutra.

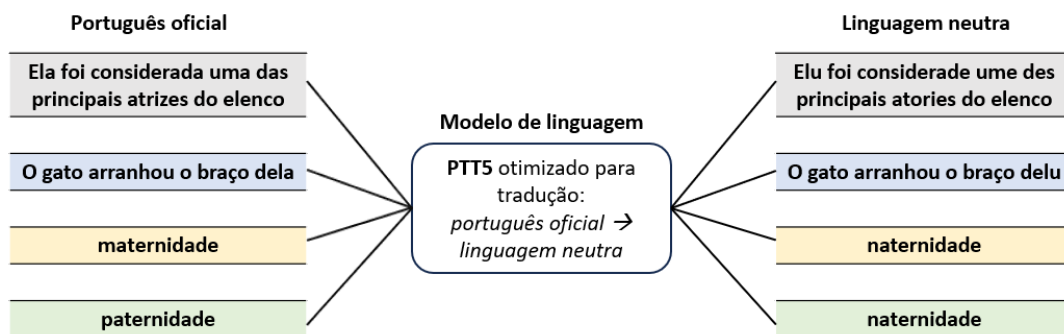


Figura 2. Diagrama instanciando o PTT5 no problema de tradução de português oficial para linguagem neutra considerando um framework do tipo texto-para-texto. Figura inspirada na apresentação do modelo T5 [Raffel et al. 2020].

2.4. Medidas de avaliação

Neste estudo, duas medidas de avaliação de tradução são aplicadas: acurácia e BLEU [Papineni et al. 2002]. Ambas são medidas supervisionadas, ou seja, medem a diferença entre a tradução realizada e a tradução de referência para a sentença traduzida.

A medida de acurácia utilizada compara as sentenças do corpus de teste com as sentenças de referência. A cada correspondência total entre tradução e referência, a acurácia é incrementada. Um valor relativo ao tamanho do corpus de teste é retornado. O resultado da medida de acurácia está dentro do intervalo $[0, 1]$ e quanto maior, melhor.

A medida BLEU (*bilingual evaluation understudy*), pressupõe que uma tradução automática é boa se ela se aproxima de uma tradução feita por um tradutor humano profissional [Papineni et al. 2002]. A ideia é usar uma média ponderada de correspondências de frases de comprimento variável traduzidas automaticamente em relação às traduções de referência. A medida compara n-gramas da sentença traduzida com n-gramas da sentença de referência e conta as correspondências (*matches*). Trata-se de uma medida do tipo “precisão” [Papineni et al. 2002]. A medida busca a proporção de palavras (no caso, unigramas) na sentença traduzida que também ocorrem nas sentenças de referência. O resultado da medida BLEU está dentro do intervalo $[0, 1]$ e quanto maior, melhor.

3. Estudos relacionados

Existem estudos que abordam a neutralização de gênero na tradução entre idiomas para analisar ou evitar a perpetuação de vieses. Em [Prates et al. 2020] é estudada a tradução de posições de trabalhos usando o Google Tradutor. Foi concluído que a plataforma tende a usar o masculino como padrão, tanto em campos com pouca representatividade feminina, ou em áreas com uma proporção igualitária de distribuição de gênero. Em [Piergentili et al. 2023] são exploradas diretrizes para uma linguagem inclusiva e os desafios técnicos ao usar modelos de tradução automática entre inglês e italiano. São definidas três orientações para quando a marcação de gênero deve ser evitada ou preferida: gênero

não deve ser expressado na saída da tradução quando não pode ser devidamente assumido na entrada; expressões de gênero apropriadas devem ser geradas na saída se elas são (indiretamente) expressadas na entrada; e genéricos masculinos não deveriam ser propagados da entrada para a saída da tradução. Em [Savoldi et al. 2021], são revisados e resumidos os trabalhos atuais na área de viés de gênero, além de discutir as estratégias de mitigação propostas até então. Como direções futuras, é mencionada a necessidade de superar a binariedade e atuar numa linguagem não-binária, remover o viés dos modelos existentes e aumentar a diversidade nas equipes profissionais de tecnologia.

Alguns estudos focam em calcular e mitigar diretamente o efeito do viés de gênero no processamento da linguagem natural. Em [Saunders et al. 2020], são propostos esquemas para incorporar marcas explícitas de flexão de gênero no nível da palavra em tradução automática implementadas com arquiteturas de redes neurais. É explorado o potencial dessa tradução controlada por flexão de gênero quando a característica de gênero pode ser determinada a partir de uma referência humana, ou quando uma sentença de teste pode ser automaticamente marcada por gênero, avaliando a tradução de inglês para espanhol e de inglês para alemão. Além disso, é proposta uma forma de avaliar a tradução neutra em termos do inglês. Em [Cho et al. 2019] é proposto um esquema para a construção de um conjunto de teste que avalia o viés de gênero em um sistema de tradução automática na língua coreana utilizando como base uma métrica chamada índice de viés de gênero na tradução (do inglês *translation gender bias index*).

Embora tenha sido realizada uma extensiva busca de literatura, esse conjunto de estudos correlatos não se apresenta como uma lista exaustiva, e sim como exemplos da preocupação internacional com a questão de marcação de gênero e vieses em uma língua. Não foi encontrado nenhum trabalho sobre tradução que lide com uma linguagem neutra baseada na inserção de novos elementos linguísticos para um idioma. O estudo [Saunders et al. 2020] é o que mais se aproxima do estudo apresentado no presente artigo, uma vez que trabalha no nível de adequação de palavras, em especial, da aplicação de pronomes. Entretanto, não lida com a inserção de novos elementos linguísticos. Assim, o presente estudo se difere das pesquisas mencionadas, pois foca na exploração de um modelo de linguagem para tradução de uma língua oficial para uma linguagem que adiciona elementos linguístico nesta mesma língua (no caso, o português).

4. Definição do problema e estudo proposto

O problema sob estudo nesta pesquisa diz respeito à tradução de textos escritos em português oficial para textos escritos em português com linguagem neutra, via aplicação de modelos de linguagens. Basicamente, um texto em português oficial é dado como entrada e seu correspondente na linguagem neutra deve ser automaticamente produzido como saída, em um modelo fim-a-fim.

A tradução pretendida possui aspectos importantes que influenciam na análise da eficácia do modelo tradutor. Em especial, para o caso tratado neste estudo, ressalta-se a necessidade de poucas modificações em uma sentença. Por um lado, há bem poucas palavras a serem traduzidas, por outro lado, não é assumida nenhuma mudança de ordem das palavras⁴. A linguagem neutra assume uma série de regras que se referem a palavras

⁴Para um parafraseador de português com objetivo de geração de uma sentença com sintaxe neutra, as características do resultado da tarefa são diferentes.

e seu uso nas sentenças. Por exemplo, “substituição dos pronomes pessoais *ela(s)* ou *ele(s)* pelo pronome neutro *elu(s)*” e “quando a palavra termina em *-a* ou *-o*, substitui a desinência por *-e*”. A aplicação de regras desse tipo sobre uma sentença em português oficial leva à necessidade de substituição de poucas palavras. Essa característica implica em particularidades a serem consideradas quando medidas de avaliação de desempenho do tradutor são aplicadas. Como exemplo, observe as instâncias de tradução apresentadas na Tabela 3. No primeiro caso, a proporção de palavras dentro da sentença que devem ser traduzidas é menor do que no segundo exemplo. A medida BLEU retorna uma avaliação mais precisa, de acordo com a tarefa em resolução, quando essa proporção é maior.

Tabela 3. Sensibilidade da medida BLEU para o problema de tradução em discussão neste estudo. # total: quantidade de palavras na sentença; # real: quantidade de palavras a ser traduzidas; # predito: quantidade de palavras traduzidas corretamente. Implementação da medida BLEU conforme descrito na Seção 5.

Tendo em vista o imenso poder **dos envolvidos** e o grau de sigilo com que **eles** operam até agora a transparência é crucial para que o Brasil tenha um entendimento claro do que **eles** realmente fizeram

Tendo em vista o imenso poder **des envolvidos** e o grau de sigilo com que **elus** operam até agora a transparência é crucial para que o Brasil tenha um entendimento claro do que **elus** realmente fizeram

# total	# real	# predito	BLEU	# predito	BLEU	# predito	BLEU
36	4	4	1,00	2	0,89	0	0,75
		3	0,92	1	0,82		

Ele criou regras para **os caçadores os atiradores** e **os colecionadores todos os** demais estão **isentos**

Elu criou regras para **es caçadories es atiradories** e **es colecionadories todes es** demais estão **isentes**

# total	# real	# predito	BLEU	# predito	BLEU	# predito	BLEU
16	10	10	1,00	6	0,60	2	0,22
		9	0,93	5	0,53	1	0,15
		8	0,74	4	0,39	0	0,11
		7	0,61	3	0,31		

Outro aspecto importante é que, de acordo com Ofélia Cassiano [Cassiano 2023], “o sistema de Linguagem Neutra, no momento e na maioria das vezes, trata do gênero de seres (em especial, pessoas)”. Animais, objetos e afins não estão suscetíveis às regras. Essa característica leva à tradução de ocorrências especiais de palavras. Por exemplo, um artigo definido que designa um animal e a desinência não serão modificados: “a gata” e “o gato” não serão traduzidos para “e gate”. Assim, em uma mesma sentença, diferentes ocorrências de um mesmo elemento linguístico poderá receber tratamentos diferentes.

Diante deste quadro de complexidades, o estudo apresentado neste artigo configura-se como uma pesquisa exploratória. O problema foi dividido em partes representadas pelos três corpora criados. Além disso, as métricas tradicionais de análise de eficiência foram aplicadas na análise dos resultados obtidos como uma forma de apresentar a sensibilidade do modelo tradutor e das medidas a cada uma das situações exploradas.

5. Método

A Figura 3 ilustra o processo de desenvolvimento e teste do modelo de linguagem inclusivo. O processo envolveu três fases que seguem detalhadas no restante desta seção.

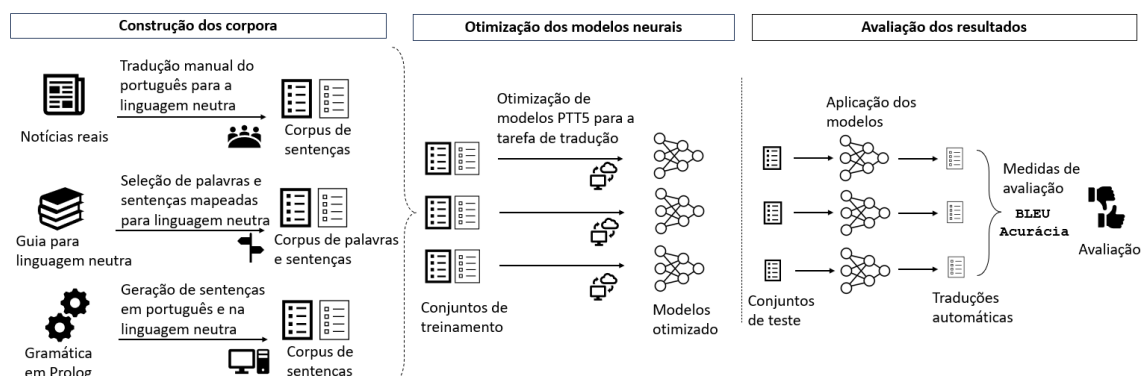


Figura 3. Visão geral do método de desenvolvimento e teste do modelo de linguagem inclusivo.

5.1. Construção dos corpora

Três corpora⁵ de naturezas diferentes foram construídos para possibilitar a exploração das capacidades dos modelos de linguagem otimizados. O primeiro corpus, baseado na tradução de trechos de notícias jornalísticas foi criado com o objetivo de otimizar um modelo a partir de aspectos do discurso real, permitindo a análise da complexidade do problema sob investigação no seu uso prático. O segundo corpus contou com palavras e sentenças usadas como exemplo no Guia para Linguagem Neutra [Cassiano 2023]. A ideia foi estudar um modelo gerado a partir de átomos e sentenças simples que ilustram o espectro da teoria da linguagem neutra de forma geral. O último corpus, baseado em uma gramática adaptada do português, foi construído a partir da geração de sentenças usando a linguagem Prolog. O objetivo com esse corpus foi otimizar um modelo que recebesse uma variedade grande de sentenças com diferentes aspectos gramaticais. Cada um dos corpora e seu processo de criação seguem descritos na continuidade desta seção.

Corpus 1: Tradução manual de parágrafos em linguagem neutra A tradução manual de parágrafos do português oficial para a linguagem neutra exigiu um trabalho minucioso. A equipe envolvida na tradução⁶ participou de palestras sobre a linguagem neutra, estudou o Guia da Linguagem Neutra [Cassiano 2023] e realizou workshops para esclarecimento de dúvidas durante o processo de tradução. A equipe selecionou parágrafos de notícias jornalísticas referentes aos temas política, cultura, esporte, segurança, educação, e legislação universitária. Para cada tema, os parágrafos foram separados em sentenças. O trabalho de tradução foi realizado aos pares. Duas pessoas foram envolvidas na tradução dos parágrafos de cada tema. Enquanto uma pessoa gerou a tradução de parte dos parágrafos, a outra fez uma revisão desta tradução, buscando melhorar sua qualidade.

⁵Disponíveis em <https://github.com/C4AI/CID>.

⁶Além dos autores deste artigo, trabalharam na construção deste corpus: Angelo Indini Anhaia, Mariluce Pereira de Jesus, Stephano Luiz da Silva Varela, Vanda Lucia Lima de Araujo, Willian Miranda Araujo da Silva e Yasmin da Silva Alexandre.

Cada dupla gerou e revisou a tradução de cerca de 100 sentenças. O corpus total de parágrafos traduzidos manualmente possui 554 sentenças. Todas as sentenças foram usadas na otimização ou no teste do modelo de linguagem. A Tabela 4 traz exemplos de sentenças pertencentes a esse corpus. Neste, todas as palavras associadas com flexão de gênero e associadas às pessoas foram traduzidas, independentemente do contexto que, eventualmente, permite a designação do gênero (veja o segundo exemplo na tabela)⁷.

Tabela 4. Exemplos de sentenças originais e traduzidas manualmente para a linguagem neutra. Cores nas palavras indicam os trechos das sentenças alvos de tradução. PT: português; LN: Linguagem neutra.

PT	Os andidatos ao mestrado e ao doutorado deverão demonstrar proficiência em, pelo menos, uma língua estrangeira, de acordo com critérios estabelecidos pela CPG.
LN	Es andidates ao mestrado e ao doutorado deverão demonstrar proficiência em, pelo menos, uma língua estrangeira, de acordo com critérios estabelecidos pela CPG.
PT	Com enorme pesar que recebemos a notícia da morte do grande intelectual, ensaísta, filósofo e diplomata Sérgio Paulo Rouanet, no último domingo, dia 3 de julho.
LN	Com enorme pesar que recebemos a notícia da morte de grande intelectual, ensaísta, filósofe e diplomata Sérgio Paulo Rouanet, no último domingo, dia 3 de julho.
PT	Quase todas as famílias de Igbo-Ora, no sudoeste da Nigéria, têm gêmeos (idênticos ou não) entre seus membros, segundo o chefe local da cidade.
LN	Quase todas as famílias de Igbo-Ora, no sudoeste da Nigéria, têm gêmies (idêntiques ou não) entre sues membros, segundo ê chefie local da cidade.

Corpus 2: Seleção de sentenças/palavras do Guia para Linguagem Neutra O Guia para Linguagem Neutra [Cassiano 2023] compreende regras e exemplos que explicam como é formada a linguagem neutra. Tais regras e exemplos dizem respeito à (mas não são restritos à) incorporação de novos elementos na língua portuguesa para pronomes, artigos, desinência e outros elementos. No guia, há exemplos de sentenças e também de palavras isoladas. Para este corpus, 656 sentenças/palavras foram selecionadas. A Tabela 5 traz exemplos pertencentes a esse corpus.

Corpus 3: Sentenças geradas via Prolog Uma gramática e um léxico para geração de sentenças em português e linguagem neutra foram desenvolvidos com linguagem Prolog. A gramática contou com os seguintes elementos linguísticos: sintagma nominal, verbal, adjetival e adverbial, e cláusulas relativas para construção da estrutura geral das sentenças; determinantes, substantivos, pronomes retos, possessivos e relativos, adjetivos,

⁷Esse procedimento foi escolhido devido ao caráter exploratório deste estudo, porém, é importante salientar que a flexão de gênero pode considerar a coexistência das três opções (“do”, “da” e “de”, no exemplo), guiada pelo contexto da sentença. A complexidade de considerar a coexistência é maior do que se espera tratar com o presente estudo, mas não deixa de ser relevante para estudos futuros e aprofundados.

Tabela 5. Exemplos de palavras e sentenças extraídas do Guia para Linguagem Neutra [Cassiano 2023]. Cores nas letras/palavras indicam os trechos das palavras e das sentenças alvos de tradução. PT: português; LN: linguagem neutra.

PT	Pintora	Nenhuma nem outra se machucaram.
LN	Pintore	Nenhume nem outre se machucaram.
PT	Sejam cuidadosos.	Ela bebeu muita água.
LN	Sejam cuidadoses.	Elu bebeu muita água.

verbos intransitivos e transitivos diretos, objetos diretos, preposições; ênclises, mesóclises e próclises. Na gramática, concordância nominal entre gênero foi usada para construir sentenças com os gêneros feminino e masculino e suas correspondentes no gênero neutro para possibilitar a construção do corpus rotulado (*sentença em português no feminino* → *sentença na linguagem neutra* e *sentença em português no masculino* → *sentença na linguagem neutra*). A concordância nominal de número foi usada para trazer diversidade para o corpus. Um léxico foi construído para instanciar a gramática. Nele foram inseridos as instâncias para determinantes e pronomes retos, exemplos para pronomes, ênclise, mesóclise e próclise, e exemplos de substantivos, verbos e adjetivos. A instanciação da gramática com o léxico gerou 21.846 sentenças. Embora esse volume de sentenças tenha sido criado, ele não foi totalmente utilizado, pois 600 pares foram amostrados para que o corpus ficasse balanceado em relação aos outros dois e para evitar frases muito similares entre si. A Tabela 6 traz exemplos de sentenças pertencentes a esse corpus, dentre aquelas que foram usadas na otimização e teste do modelo. Neste corpus, embora as sentenças sejam gramaticalmente corretas, elas não são, necessariamente, corretas em termos pragmáticos (veja o terceiro exemplo na tabela) e, além disso, uma única sentença não contém os gêneros masculino e feminino combinados no discurso.

Tabela 6. Exemplos de sentenças sintéticas geradas com a linguagem Prolog. Cores nas palavras indicam os trechos das sentenças alvos de tradução. PT: português; LN: linguagem neutra.

PT feminino	a menina correrá
LN	ê menine correrá
PT masculino	ele ensina-los-a como nadar
LN	elu ensina-les-a como nadar
PT feminino	as gatas as quais elas admiram veem as tuas gatas lindas
LN	as gatas as quais elus admiram veem as tuas gatas lindas
PT masculino	os gatos os quais eles admiram veem os teus gatos lindos
LN	os gatos os quais elus admiram veem os teus gatos lindos

5.2. Otimização e avaliação do modelo tradutor

A implementação⁸ do modelo foi feita em linguagem Python. O código utilizado foi originalmente desenvolvido para uma tarefa de geração de texto, e pequenas modificações foram feitas para adapta-lo à tarefa de tradução. Nessa implementação, o tokenizador é otimizado usando o “ptt5-base-portuguese-vocab” [Chen and Cherry 2014], a otimização

⁸<https://github.com/ThiagoCF05/Any2Some>

é realizada a partir do “checkpoint” do PTT5⁹, usando a função de perda entropia-cruzada e o otimizador AdamW¹⁰. A otimização do modelo foi feita com uma estratégia de treinamento com parada antecipada. Os corpora foram divididos em três subconjuntos disjuntos de pares de sentenças (e pares de palavras no caso do corpus 2): 80% para treino, 10% para validação e 10% para teste. Um estudo exploratório foi realizado para os hiperparâmetros taxa de aprendizado, tamanho de lotes e comprimento máximo de *tokens*, sendo os valores finais, 10^{-3} , 16 e 250, respectivamente. Para avaliação do modelo tradutor, a implementação da acurácia segue a lógica descrita na Seção 2 e para a medida BLEU, um método de suavização foi utilizado¹¹ [Chen and Cherry 2014]. O treinamento e o teste foram feitos combinando, de diferentes formas, os subconjuntos de sentenças (e palavras) de treinamento, de validação e de teste (cf. Seção 6).

6. Resultados e discussões

A Tabela 7 apresenta a variabilidade linguística de cada corpus: a quantidade de *tokens*, a quantidade de *types* (unidades linguísticas sem repetição), e a razão entre elas. Essas informações esclarecem a dificuldade trazida pelo vocabulário em cada um dos corpus.

Tabela 7. Variabilidade linguística de cada corpus em termos de *tokens*, *types* e a relação entre eles. #: quantidade; PT: português; LN: linguagem neutra.

		# de Tokens	# de Types	Types/Tokens
<i>Corpus 1</i>	PT	104.950	5.426	5,17%
	LN	108.132	5.269	4,87%
<i>Corpus 2</i>	PT	71.823	2.254	3,14%
	LN	72.448	2.068	2,85%
<i>Corpus 3</i>	PT	76.533	1.456	1,90%
	LN	79.465	1.439	1,81%

O primeiro teste foi aplicar o modelo de linguagem sem otimização (Tabela 8 - *Baseline*), criando uma base para observação do desempenho do modelo. Essa verificação foi realizada usando o conjunto total de sentenças de cada corpus e o subconjunto de teste de cada um deles. Como esperado, a acurácia “0” foi obtida para todos os casos, já que todas as sentenças para avaliação estão na linguagem neutra e possuem pelo menos uma palavra do conjunto de novos elementos linguísticos. A medida BLEU exige uma análise mais refinada, mas algumas hipóteses podem ser levantadas de acordo com o discutido na Seção 4: no *corpus 1* as sentenças são longas e no *corpus 2* elas são curtas, porém ambas possuem poucas palavras a serem traduzidas e a medida BLEU mostra a faixa de avaliação obtida, com um viés em favor da quantidade de palavras que não precisaram ser traduzidas; no *corpus 2* há instâncias com apenas uma palavra e ela deve ser traduzida e, o método de suavização para cálculo do BLEU usado é sensível a essas situações¹². O segundo teste é um teste de ressubstituição [Han and Kamber 2006], ou seja, o conjunto

⁹<https://huggingface.co/unicamp-dl/ptt5-base-portuguese-vocab>

¹⁰<https://pytorch.org/docs/stable/generated/torch.optim.AdamW.html>.

¹¹Método implementado na função `smoothing_function=chencherry.method3` disponível em https://www.nltk.org/api/nltk.translate.bleu_score.html.

¹²Há um outro método que poder ser utilizado nesses casos (`method4`), mas ele é sensível à situação inversa. Optamos, neste estudo em considerar o problema de sentenças maiores na avaliação via BLEU.

usado no treinamento é usado no teste, obtendo-se assim uma avaliação do desempenho de aprendizado do modelo. Os resultados obtidos (Tabela 8 - Resubstituição) mostram que houve um aprendizado para todos os casos, com uma expressividade maior para o caso do *corpus 3*.

Tabela 8. Resultados: sem treinamento (baseline); no teste de ressubstituição.

	<i>Baseline</i> (100 - 10% das sentenças)			Desempenho de ressubstituição		
	<i>Corpus 1</i>	<i>Corpus 2</i>	<i>Corpus 3</i>	<i>Corpus 1</i>	<i>Corpus 2</i>	<i>Corpus 3</i>
BLEU	0,72 - 0,73	0,11 - 0,11	0,37 - 0,38	0,86	0,30	0,99
Acurácia	0,00 - 0,00	0,00 - 0,00	0,00 - 0,00	0,30	0,36	1,00

Na Tabela 9 estão listados os resultados considerando: treinamento e teste no mesmo corpus (em negrito na tabela); treinamento em um corpus e teste no outro; treinamento com todos os corpora e teste também com todos os corpora (última linha da tabela). Os conjuntos de treinamento e teste são disjuntos, como descrito na Seção 5.

Tabela 9. Resultados: no conjunto de teste, considerando análises cruzadas.

	<i>Corpus 1</i>		<i>Corpus 2</i>		<i>Corpus 3</i>		Todos (teste)	
	BLEU	Acurácia	BLEU	Acurácia	BLEU	Acurácia	BLEU	Acurácia
<i>Corpus 1</i>	0,88	0,30	0,31	0,10	0,30	0,08	0,73	0,11
<i>Corpus 2</i>	0,77	0,01	0,28	0,31	0,39	0,01	0,64	0,21
<i>Corpus 3</i>	0,73	0,05	0,14	0,66	0,99	1,00	0,70	0,33
Todos (treino)	0,84	0,20	0,45	0,61	0,99	0,98	0,84	0,61

Como já observado, a compreensão da medida BLEU requer análises aprofundadas sobre a relação entre palavras a serem traduzidas e tamanho de sentenças. Contudo, o que se observa na Tabela 9 é que há um ganho no aprendizado do modelo após a otimização quando comparado ao *baseline* e em relação a todos os testes com o *corpus 1* e *corpus 2*. Para o *corpus 3* observa-se que, quando usado como teste para treinamentos feitos com o *corpus 1* e o *corpus 2*, não acusa qualidade de aprendizado. Entretanto, quando o treinamento envolve o *corpus 3*, a medida BLEU acusa alta qualidade de aprendizado no teste para o mesmo corpus. A medida de acurácia acusa bons resultados principalmente em relação ao treinamento com o *corpus 3* e teste com o *corpus 2* e *corpus 3*. Esses dois corpora são fortemente caracterizados por sentenças curtas, e particularmente o *corpus 3* possui muitos exemplos com traduções nas terminações das palavras. O *corpus 1*, que possui sentenças longas se mostra mais complexo para o modelo de linguagem.

Como expressão da factibilidade de uso do modelo de linguagem para a tarefa de tradução em questão neste estudo, observa-se o aprendizado obtido quando todos os subconjuntos são envolvidos no treino ou no teste. A medida BLEU, neste caso, traz um resultado médio, já que está aplicada a conjuntos de sentenças de tamanhos variados. A medida de acurácia traz um resultado bastante interessante, uma vez que deve-se considerar que os corpora usados para otimização do modelo não são tão grandes quanto poderiam (ou deveriam) ser, para fornecer um ambiente mais suscetível para promoção do aprendizado de um modelo de linguagem.

7. Considerações Finais

Este artigo descreve um estudo que pode ser colocado como a primeira iniciativa de exploração de um modelo de linguagem para tradução do português oficial para a linguagem neutra, considerando novos elementos linguísticos para neutralização de gênero. Os resultados, embora ainda preliminares, representam um marco para a discussão da construção de modelos de linguagem inclusivos. A proposição e disponibilização dos corpora e a discussão realizada abre caminhos para estudos mais aprofundados.

Entre os esforços futuros pertinentes neste contexto está o estudo de grandes modelos de linguagem recentemente criados. Um teste rápido usando o sistema ChatGPT¹³, sob o paradigma “*pre-train, prompt, and predict*” [Liu et al. 2023], para uso de sintaxe neutra mostra que, embora o desempenho do modelo seja bom, ele não atende ao que é necessário para uma linguagem inclusiva. O ChatGPT tem um bom desempenho para o *prompt* “*Escreva as frases seguintes usando linguagem neutra: <[frases]>*”, mas usa alguns artifícios inadequados. Exemplos positivos desse uso são as seguintes reescritas: “*Aqueles que terminaram podem sair* → *Quem terminou pode sair*” e “*Nós somos alunos de teatro* → *Nós somos estudantes de teatro*”. Exemplos negativos são: “*As cantoras denunciaram os assédios* → *As artistas denunciaram os assédios*, no qual o gênero feminino não foi neutralizado, e “*Você é biólogo?* → *Você é biólogx?*, no qual o uso do caractere “*x*” como neutralizador foi aplicado. Em relação ao uso de novos elementos linguísticos, como explorado no presente estudo, o potencial de sucesso com um ChatGPT é factível, mas necessita de um trabalho de engenharia de *prompt* ou de otimização do modelo mais aprofundado. Um *prompt* simples no estilo “*few-shot learning*” ou um *prompt* indicando qual é o novo elemento linguístico e sua classe gramatical são suficientes para que o sistema faça traduções corretas desse elementos (por exemplo: “*Eles são docentes* → *Elus são docentes*”. Porém, tentativas de fazer o sistema entender regras de desinência aplicadas apenas no que diz respeito à pessoas não geraram bons resultados (por exemplo: “*Tendo em vista o imenso poder deles ...* → *tende em viste o imense podere delus ...*”).

Este estudo manifesta apoio à discussão sobre neutralização de gênero na língua portuguesa, mas não advoga que a linguagem neutra, tal como proposta, é apropriada em todos os seus aspectos e para todos os efeitos sociais e linguísticos que sua adoção pode produzir. O estudo coloca a discussão sob o olhar da inteligência artificial, e situa essa disciplina como agente conhecedor das nuances da sociedade e da importância de ferramentas que atendam todos os grupos sociais e iniciativas focadas no bem social.

8. Agradecimentos

Este trabalho foi executado no Centro de Inteligência Artificial (C4AI-USP) com apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation. Os autores agradecem a colaboração de Thiago Castro Ferreira e João Gabriel Moura Campos.

Referências

Carmo, D., Piau, M., Campiotti, I., Nogueira, R., and Lotufo, R. (2020). PTT5: Pre-training and validating the T5 model on brazilian portuguese data. *arXiv preprint arXiv:2008.09144*.

¹³<https://chat.openai.com/>: criado a partir da otimização de um modelo da série GPT-3.5

- Cassiano, O. (2023). Guia para “linguagem neutra” (PT-BR). Online. Acessado em 01/07/2023.
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., and Xie, X. (2023). A survey on evaluation of large language models.
- Chen, B. and Cherry, C. (2014). A systematic comparison of smoothing techniques for sentence-level BLEU. In *Proc. of the 9th Workshop on Stat. Mach. Transl.*, pages 362–367, Baltimore, Maryland, USA. ACL.
- Cho, W. I., Kim, J. W., Kim, S. M., and Kim, N. S. (2019). On measuring gender bias in translation of gender-neutral pronouns. In *Proc. of the 1st Workshop on Gender Bias in Nat. Lang. Process.*, pages 173–181, Florence, Italy. ACL.
- Han, J. and Kamber, M. (2006). *Data Mining. Concepts and Techniques*. Morgan Kaufmann, 2nd ed. edition.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.*, 55(9).
- Mäder, G. R. C. (2015). Masculino genérico e sexismo gramatical. Master’s thesis, Universidade Federal de Santa Catarina.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proc. of the 40th Annual Meeting on ACL*, pages 311–318. ACL.
- Piergentili, A., Fucci, D., Savoldi, B., Bentivogli, L., and Negri, M. (2023). From inclusive language to gender-neutral machine translation. *arXiv preprint arXiv:2301.10075*.
- Prates, M. O., Avelar, P. H., and Lamb, L. C. (2020). Assessing gender bias in machine translation: a case study with google translate. *Neural Comput. Appl.*, 32:6363–6381.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1).
- Saunders, D., Sallis, R., and Byrne, B. (2020). Neural machine translation doesn’t translate gender coreference right unless you make it. In *Proc. of the 2nd Workshop on Gender Bias in Nat. Lang. Process.*, pages 35–43, Barcelona, Spain (Online). ACL.
- Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., and Turchi, M. (2021). Gender bias in machine translation. *Trans. of ACL*, 9:845–874.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Inf. Process. Syst.*, volume 30. Curran Associates, Inc.
- Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brWaC corpus: A new open resource for Brazilian Portuguese. In *Proc. of the 11th Int. Conf. on Lang. Resour. Eval. (LREC 2018)*, Miyazaki, Japan. ELRA.