# Uma Investigação sobre Técnicas de Data Augmentation Aplicadas a Tradução Automática Português-LIBRAS

Marcos André Bezerra da Silva
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
marcos.andre@lavid.ufpb.br

Manuella Aschoff C. B. Lima
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
manuella.lima@lavid.ufpb.br

Diego Ramon Bezerra da Silva
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
diego.silva@lavid.ufpb.br

Daniel Faustino L. de Souza
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
daniel@dcx.ufpb.br

Rostand Edson O. Costa
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
rostand@lavid.ufpb.br

Tiago Maritan U. de Araújo
Universidade Federal da Paraíba
João Pessoa/PB, Brasil
tiagomaritan@lavid.ufpb.br

## ABSTRACT

The automatic translation from Portuguese to LIBRAS is extremely important for accessibility and inclusion of deaf individuals in society, but the scarcity of data and the high cost of building an authentic corpora pose significant challenges. Data Augmentation in Neural Machine Translation is the process of generating synthetic sentences to increase the quantity and diversity of the training set. This work investigates the use of data augmentation techniques to improve the performance of Portuguese-LIBRAS automatic translation using the BLEU metric. Among the techniques analyzed, back-translation and its combination with synonym substitution using part-of-speech tagging stood out as the most effective in enhancing the translation model and can be used to increase the diversity of underrepresented datasets.

## KEYWORDS

Tradução Automática Neural, Aumento de Dados, Libras

## 1 INTRODUÇÃO

A evolução da tecnologia tem desempenhado um papel crucial na acessibilidade e inclusão de pessoas surdas, representando um marco significativo na promoção da igualdade de acesso à informação. Uma parcela considerável da população surda não compreende os textos disponibilizados na forma escrita da língua oral (LO), comunicando-se, portanto, basicamente através da língua de sinais (LS). Especialmente na *web*, onde o volume e o dinamismo de informações são enormes, com conteúdo textual sendo gerado a todo instante, a tarefa de interpretar manualmente textos de páginas *web* para língua de sinais é inviável. Diante desse cenário, torna-se indispensável o uso de componentes de tradução automática para traduzir o conteúdo, por exemplo, do Português Brasileiro (PB) para a Língua Brasileira de Sinais (LIBRAS), para que pessoas surdas tenham acesso efetivo à informação online [17]. Nesse contexto, a plataforma VLibras [1] emerge como uma solução que torna a *web* verdadeiramente acessível para surdos, destacando-se por sua ampla adoção em sites governamentais, onde desempenha um papel essencial ao prover acessibilidade em LIBRAS para os serviços públicos. Atualmente, o VLibras está sendo utilizado em mais de 500.000 sites, tanto públicos quanto privados, realizando milhões de traduções mensalmente [2].

O componente de tradução do VLibras foi inicialmente desenvolvido utilizando regras de tradução advindas da expertise de linguistas. Na abordagem atual do VLibras, é utilizado um tradutor baseado em Tradução Automática Neural, que é capaz de inferir as regras da linguagem a partir dos dados de treinamento. A adição do componente baseado em rede neural possibilitou uma maior capacidade de desambiguação de palavras em PB que possuem sinais diferentes na LIBRAS, a depender do contexto [17].

Entretanto, apesar dos avanços trazidos pelo uso de redes neurais para Tradução Automática, é exigida uma alta quantidade de dados para o treinamento de modelos que geram traduções de boa qualidade [14], o que é um grande obstáculo, principalmente em LS, como a LIBRAS. Para que isso aconteça é necessário o desenvolvimento de um corpus bilíngue[1], processo que é extremamente custoso e manual, visto que é feito por intérpretes da LIBRAS. Sendo assim, é um desafio construir um conjunto de dados de treinamento diverso que represente diferentes contextos de uso da língua [17]. Neste sentido, a geração de dados sintéticos por meio de técnicas de aumento de dados (*data augmentation*) é uma estratégia importante para superar a escassez e aumentar a quantidade e diversidade dos dados de treinamento [19].

Existem vários métodos de aumento de dados para Processamento de Linguagem Natural (PLN) e muitos métodos, apesar de poderem ser utilizados em diversas tarefas, foram elaborados com a finalidade de melhorar o desempenho em uma tarefa específica. A escolha certa dos métodos empregados para aumento de dados pode trazer um impacto positivo na qualidade dos modelos gerados, visto que é sabido que métodos projetados para a tarefa de Tradução Automática para línguas com muitos recursos disponíveis podem não ser eficazes para línguas com pouco recurso (*low-resource languages*) [4]. Sendo assim, este trabalho tem como objetivo investigar o impacto de diferentes métodos de aumento de dados, utilizando técnicas como retrotradução, substituição de palavras alinhadas, reversão de *tokens* e substituição por sinônimos com *part-of-speech tagging*, além de analisar a influência da quantidade de dados sintéticos no desempenho do modelo de tradução PB-LIBRAS. Para isso, foram realizados experimentos com um modelo da arquitetura

[1] Pares de sentenças em português e suas respectivas traduções em LIBRAS.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2024). Juiz de Fora, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2966-2753

*transformer* e os resultados foram avaliados com base na métrica BLEU.

## 2 TRABALHOS RELACIONADOS

A tradução automática PB-LIBRAS envolve a tradução entre línguas de modalidades diferentes (oral-auditiva e visual-espacial) e que possuem estruturas linguísticas não paralelas entre si. Além disso, as LS geralmente têm poucos recursos, ou seja, existem poucos bancos de dados extensos para LS e, quando existem, são limitados a algumas LS específicas [5, 10, 15]. Assim, a escassez de dados é um dos principais desafios para a tradução automática para língua de sinais (*Sign Language Translation* - SLT). Algumas estratégias para lidar com o problema de tradução automática com poucos recursos incluem aumento de dados, aprendizagem por transferência, retrotradução e abordagens híbridas [8]. No tocante à técnica de aumento de dados, foco deste trabalho, destacam-se os trabalhos de Moryossef et al. 2021, Fadaee et al. 2017, Sanchez-Cartagena et al. 2021, Maimaiti et al. 2021, Jang et al. 2022 e Wang, Yang 2022.

Moryossef et al. 2021 se concentraram na tarefa de tradução de texto dentro do SLT e introduziram duas estratégias de aumento de dados baseadas em regras. Eles apresentaram regras abrangentes e específicas do idioma para criar pares texto-glosa pseudo-paralelos. Esses pares foram posteriormente empregados no processo de retrotradução, melhorando o desempenho geral do modelo.

O trabalho desenvolvido por Fadaee et al. 2017 utilizou uma estratégia de substituição de palavras alinhadas para aumentar a frequência de palavras raras no conjunto de treinamento. As sentenças geradas foram posteriormente filtradas por um modelo de linguagem que foi treinado com o objetivo de avaliar se as sentenças sintéticas são fluentes, ou seja, estão corretas gramaticalmente e fazem sentido semântico. Foi observado um ganho de 2,5 pontos na métrica BLEU na tradução de inglês para alemão.

Já Sánchez-Cartagena et al. 2021 aplicaram duas técnicas para aumento de dados em sua pesquisa. Uma delas é a substituição de palavras alinhadas semelhante à apresentada por Fadaee et al. 2017, porém sem se preocupar com a fluência das sentenças geradas, portanto, não sendo necessário um modelo de linguagem adicional. De maneira auxiliar, também foi utilizada a tarefa de reversão de *tokens*, tendo os resultados avaliados na tradução entre inglês e alemão, hebreu e vietnamita, resultando em um ganho médio de 1,6 BLEU.

Maimaiti et al. 2021 em sua pesquisa propuseram a substituição de sinônimos com *part-of-speech tagging*. Esse trabalho aplicou essa técnica para tradução dos idiomas azerbaijão, hindi, uzbeque, turco, alemão e chinês para inglês. Como resultado, foram observados ganhos de BLEU entre 1,16 e 2,39 pontos.

Na pesquisa realizada por Jang et al. 2022, foram utilizadas técnicas de aumento de dados para língua de sinais coreana (*Gloss-level Korean Sign Language*). A proposta utilizou retrotradução, substituição por sinônimos restrita às classes de substantivos, nomes próprios e pronomes, além de substituição de palavras utilizando um modelo de linguagem coreano. Foram observados ganhos de BLEU em 10, 12 e 16 pontos, respectivamente.

Por fim, destaca-se a pesquisa de Wang, Yang 2022 que trabalharam em modelos de tradução entre os idiomas inglês, chinês e tailandês. Neste trabalho foram propostos aumento de dados por substituição de palavras alinhadas e substituição por sinônimos. Observaram ganhos de 1 BLEU ao utilizar substituição de palavras alinhadas na tradução de chinês para inglês e de 4 BLEU na tradução de inglês para chinês. Já a substituição por sinônimos na tradução de chinês para tailandês resultou em um ganho de 2,7 BLEU.

## 3 FUNDAMENTAÇÃO TEÓRICA

### 3.1 Tradução Automática Neural

Tradução Automática Neural (*Neural Machine Translation* - NMT) é a aplicação de redes neurais para a tarefa de tradução de sentenças de uma língua de origem para uma língua de destino. Em geral, é utilizada uma rede *sequence-to-sequence*, onde o *encoder* constrói uma representação da sentença no idioma de origem e o *decoder* utiliza essa representação e as palavras geradas anteriormente pela própria rede para gerar a sentença traduzida no idioma alvo [3].

Em contraste à Tradução Automática Baseada em Regras (*Rule Based Machine Translation* - RBMT), que transforma a sentença de origem através de algoritmos de substituição por regras, derivadas do conhecimento de linguistas, a NMT é baseada em dados. Sendo assim, para que a NMT seja possível, é necessária a construção de um corpus bilíngue. As redes neurais treinadas para resolver o problema de tradução são capazes de aprender as regras de tradução diretamente dos dados, eliminando assim a necessidade da construção de algoritmos explícitos com regras de tradução. Especificamente no contexto da LIBRAS, modelos NMT oferecem uma capacidade de desambiguação de palavras que têm grafia igual no PB, porém significado e sinal diferentes na LIBRAS. Essa capacidade de desambiguação não seria alcançada utilizando apenas uma abordagem de RBMT, já que é necessário que o modelo de tradução compreenda o contexto em que o termo ambíguo está inserido para decidir a desambiguação correta [17].

### 3.2 Data Augmentation

A qualidade de um modelo de Tradução Automática depende principalmente da existência de um corpus bilíngue de alta qualidade e extensão significativa. Este corpus serve como base de treinamento para que o modelo de tradução neural aprenda os padrões linguísticos de ambos idiomas de origem e destino [17]. A LIBRAS, assim como outras LS, pode ser classificada como língua com poucos recursos (*low-resource language*), tendo em vista a baixíssima quantidade de dados disponíveis para o treinamento de componentes de PLN [2]. Para as línguas com poucos recursos, não há dados autênticos traduzidos por humanos suficientes disponíveis para treinar um modelo de NMT e obter resultados de alta qualidade. Desta forma, a geração de dados sintéticos é uma estratégia interessante para complementar o corpus criado pelos linguistas.

Com um corpus pequeno, o universo de sentenças a serem traduzidas pelo modelo (quando o tradutor estiver sendo utilizado pelo usuário) será muito maior que os exemplos vistos no treinamento. O objetivo de algoritmos de aumento de dados é, a partir dos dados autênticos, expandir e diversificar o corpus de treinamento com dados sintéticos para que, idealmente, se aproxime da distribuição de dados de todo o universo de pares de sentenças e traduções válidas, de modo que a quantidade e diversidade desses novos dados beneficiem o modelo de tradução [19].

Algoritmos de aumento de dados (*data augmentation*) geram dados adicionais, sintéticos, a partir dos dados autênticos, através de modificações sobre as sentenças autênticas. Esses dados adicionais são então incorporados ao corpus de treinamento original. Esta é uma tarefa desafiadora no ramo de PLN e tradução automática, onde qualquer modificação na sentença pode ter impacto no seu significado. Por isso, é importante que as traduções na língua de destino mantenham equivalência com o significado na língua de origem. É uma alternativa mais barata na falta de exemplos curados por linguistas e é extremamente importante em línguas com poucos recursos. Porém, por ser um procedimento automático, as sentenças sintéticas geradas pelos algoritmos de aumento de dados tendem a ser de menor qualidade em relação às sentenças autênticas produzidas por humanos. Deve-se, portanto, utilizar uma quantidade razoável de dados sintéticos [14].

Existem diversas técnicas para *data augmentation* e, neste trabalho, abordaremos as técnicas de retrotradução, reversão e substituição de palavras, as quais serão detalhadas a seguir.

*3.2.1 Retrotradução.* A técnica de retrotradução (tradução reversa ou *back-translation*) é amplamente utilizada para aumento de dados, pois tende a gerar sentenças de boa qualidade. Nessa abordagem, como pode ser observado na Figura 1, é utilizado um outro modelo de tradução e a sentença é traduzida de um idioma para outro e depois de volta para o idioma original[2]. Assim, a partir de um par autêntico de uma sentença na língua de origem e sua respectiva tradução, é possível gerar novas sentenças na língua de origem que tenham o mesmo significado da tradução da sentença original. Isso é feito ao executar o algoritmo nas sentenças do idioma de origem, resultando em novas sentenças sintéticas no mesmo idioma. Essas novas sentenças, resultantes da tradução reversa, são adicionadas ao corpus original no lado do idioma de origem. As traduções correspondentes, na língua de destino, associadas a essas novas sentenças sintéticas, serão idênticas às traduções das sentenças autênticas originais devido à tendência de preservação do significado [14].

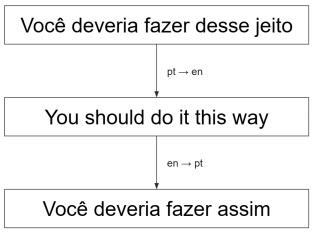


**Figura 1: Exemplo de retrotradução.**

*3.2.2 Reversão.* A reversão dos *tokens* da sentença na língua de destino, como exemplificado na Tabela 1, é uma tarefa auxiliar e não convencional. No entanto, faz com que o modelo de tradução utilize mais informações da representação do *encoder* para prever palavras que geralmente aparecem no final da frase, onde a influência do *encoder* tende a diminuir. Dado que as sentenças geradas não são fluentes, é recomendada a adição de um *token* especial no início de cada par de sentenças sintéticas [16].

[2]Por exemplo, traduzir uma sentença do português para o inglês e, em seguida, de volta para o português.

*3.2.3 Substituição de Palavras.* Técnicas baseadas em substituição de palavras envolvem gerar novas sentenças sintéticas escolhendo uma palavra alvo na sentença autêntica para ser substituída aleatoriamente por outra palavra. Isso pode ser aplicado tanto nas sentenças da língua de origem quanto nas da língua de destino, e as substituições podem ser realizadas independentemente entre si. Alternativamente, pode-se respeitar o alinhamento entre as palavras e realizar a substituição aos pares, como exemplificado na Tabela 1: ao substituir uma palavra na sentença de origem, também é substituída a palavra correspondente na sentença de destino. Mesmo que a substituição seja aleatória, a introdução de ruído nas sentenças de destino ajuda o modelo a aprender a prever a próxima palavra corretamente, mesmo após a geração de uma palavra que não corresponde exatamente ao padrão ouro[3] da tradução humana [19] [3].

Substituir palavras por sinônimos é uma abordagem que gera sentenças sintéticas com significado mais próximo da sentença autêntica. Ao escolher uma palavra para ser substituída, é consultada uma tabela de paráfrases que contém sinônimos que são candidatos a substituir a palavra original. Através da representação vetorial da palavra original e das palavras candidatas, é calculada a similaridade cosseno a fim selecionar a palavra candidata com maior similaridade para substituir a palavra original. Também é possível utilizar marcação de partes do discurso (*part-of-speech tagging*) para limitar as opções de substituição dentro da mesma classe gramatical, como exemplificado na Figura 2. Ao identificar cada classe gramatical presente em uma sentença por meio de *part-of-speech tagging*, torna-se viável realizar substituições de palavras mantendo a coerência gramatical do texto. Essas restrições também evitam erros que podem ocorrer com o uso da retrotradução, que confia totalmente na saída do modelo de tradução [11].

**Tabela 1: Exemplos de reversão e substituição alinhada.**

| Transformação | Idioma | Sentença |
|---|---|---|
| Nenhuma (sentença original) | Origem | Tivemos uma calorosa recepção. |
| | Destino | TER CALOROSO&ANIMADO RECEPÇÃO [PONTO] |
| Reversão | Destino | [PONTO] RECEPÇÃO ANIMADO&CALOROSO TER |
| Substituição alinhada | Origem | Tivemos uma calorosa *rio*. |
| | Destino | TER CALOROSO&ANIMADO *RIO* [PONTO] |

# 4 METODOLOGIA

## 4.1 Técnicas Selecionadas

Com o objetivo de identificar quais métodos de aumento de dados trazem o melhor ganho de desempenho ao tradutor, foram selecionadas as seguintes técnicas: (a) retrotradução, por ser amplamente utilizada em NMT, sendo uma referência sólida para comparação, (b) a substituição alinhada e a tarefa auxiliar de reversão como descrito por Sánchez-Cartagena et al. 2021 e exemplificado na Tabela

[3]Tradução realizada por humanos e tomada como referência.


**Figura 2: Exemplos de substituição por sinônimos com *part-of-speech tagging*. [11]**

1, dado que é uma evolução sobre métodos de substituição aleatória anteriores, sendo selecionada por sua capacidade de melhorar a diversidade dos dados sem a necessidade de modelos de linguagem adicionais e (c) a substituição por sinônimos com *part-of-speech tagging* como proposto por Maimaiti et al. 2021 e observado na Figura 2, por ser uma alternativa à retrotradução que também tende a gerar sentenças fluentes.

Para a retrotradução, são utilizados dois modelos de tradução automática disponíveis no *framework Hugging Face Transformers*: o modelo *Helsinki-NLP/opus-mt-ROMANCE-en*, que traduz do PB para o inglês, e o *modelo Helsinki-NLP/opus-mt-en-ROMANCE*, que traduz do inglês de volta para o PB. Durante a etapa de retrotradução, as sentenças em PB do corpus autêntico são traduzidas para o inglês pelo primeiro modelo e, em seguida, traduzidas de volta para o PB pelo segundo modelo. Após remover as duplicatas, o corpus expandido apresenta um aumento de cerca de 50% no número total de sentenças.

A substituição de palavras alinhadas, como apresentada por Sánchez-Cartagena et al. 2021, foi realizada utilizando o sistema de tradução automática estatística MOSES[4] que utiliza a biblioteca GIZA para o alinhamento de palavras. O MOSES produz um léxico que contém entradas de palavras na língua de origem, juntamente com a probabilidade de que a palavra correspondente na língua de destino seja uma tradução apropriada. Para a substituição, foi determinado substituir uma palavra por sentença que é sorteada aleatoriamente, devendo o par de palavras alinhadas ter uma probabilidade maior que 0.7 no léxico. Em seguida, este par de palavras é substituído por outro par de palavras, também de probabilidade maior que 0.7, proporcionando uma variação semântica na sentença.

O método de reversão consiste em inverter a lista dos *tokens* da *string* da sentença original, separados pelo caractere de espaço. Para mitigar o impacto de sentenças potencialmente não fluentes, foi adicionado um *token* especial precedendo a sentença de origem em cada método. Essa medida tem como objetivo diminuir a influência dessas sentenças no resultado final do tradutor.

[4]O MOSES é um sistema de tradução automática que opera através do alinhamento um para um de todos os *tokens* da língua de origem para a língua de destino.

A implementação da substituição por sinônimos baseada em *part-of-speech tagging*, conforme descrito por Maimaiti et al. 2021, utiliza uma combinação de técnicas e recursos de PLN, incluindo modelos de *part-of-speech tagging*, a base de dados *WordNet* para identificar sinônimos candidatos e *word embeddings* para calcular a similaridade entre os sinônimos candidatos e escolher o sinônimo com maior similaridade cosseno. Para realizar a marcação de partes do discurso, foi utilizado o modelo *POS_tagger_brill.pkl*, disponível no repositório do *GitHub inoueMashuu/POS-tagger-portuguese-nltk* para o *framework* de PLN *nltk*. Esse modelo é responsável por atribuir classes gramaticais às palavras em PB para a identificação das palavras alvo que serão substituídas por sinônimos. A base de dados *WordNet* [5], acessada através da biblioteca *nltk*, foi empregada como fonte de sinônimos para as palavras identificadas pelo *part-of-speech tagging*. Por fim, para calcular a similaridade entre as palavras alvo e os sinônimos disponíveis no *WordNet*, foram utilizadas as *embeddings skip-gram* de 100 dimensões do Repositório de *Word Embeddings* do NILC [6]. As *word embeddings* são representações vetoriais das palavras que capturam suas relações semânticas com base em seu contexto de ocorrência. Essas representações vetoriais permitem calcular a similaridade cosseno entre palavras, necessário para identificar o sinônimo mais adequado para a substituição.

## 4.2 Ambiente de Treinamento

O treinamento foi realizado utilizando o *framework Fairseq*[6]. Optou-se por uma versão reduzida de um modelo que adota a arquitetura *transformer*, proposta por Vaswani et al. (2017). No *Fairseq*, o modelo *transformer* reduzido *transformer_iwslt_de_en* foi pré-treinado na tradução de alemão para inglês. A escolha desse modelo permite um treinamento mais rápido, uma vez que fazer o ajuste fino (*fine-tuning*) de um modelo de linguagem já treinado, mesmo que em um par de idiomas diferentes, tende a ser mais eficiente do que treinar um modelo do zero [15].

O tradutor do VLibras possui uma arquitetura híbrida baseada em RBMT e NMT. A sentença em PB é pré-processada por um componente RBMT. A saída desse pré-processamento alimenta o modelo *transformer*, que é treinado com o objetivo de aproximar a glosa gerada pelo componente RBMT para a glosa gerada pelos intérpretes humanos [2].

O corpus do VLibras possui pares de sentenças em PB como língua de origem e uma representação intermediária da LIBRAS, denominada glosa, para as sentenças no idioma de destino. O corpus mencionado consiste em aproximadamente 65.000 exemplos de pares de sentenças [2], cobrindo uma ampla variedade de tópicos e contextos linguísticos. As glosas são criadas por intérpretes e linguistas especializados na LS e cada sinal da glosa possui uma animação associada. O uso de glosas intermediárias como uma forma de representação linguística da LS permite que os algoritmos de PLN trabalhem de forma mais eficaz com a LIBRAS [9].

[5]O *WordNet* é uma vasta base de dados lexical que organiza palavras em conjuntos de sinônimos, conhecidos como synsets, e fornece informações sobre suas relações semânticas e gramaticais. Essa base de dados permite a consulta de sinônimos restritos a classes gramaticais específicas, como sujeitos, adjetivos, advérbios e verbos, permitindo que a substituição gere sentenças de maior qualidade.

[6]Desenvolvido pelo *Facebook*, projetado com foco no treinamento de redes *sequence-to-sequence*, que são adequadas para tarefas de tradução automática.

No *pipeline* de tradução do VLibras, já existem cinco métodos de aumento de dados que geram sentenças fluentes e foram construídos com o conhecimento de linguistas da LIBRAS.

- **Lugares:** tem como objetivo introduzir variação nas sentenças ao identificar nomes de lugares como cidades, estados ou países, e substituí-los por outras localidades disponíveis em tabelas de substituição auxiliares.
- **Negação:** trabalha na identificação de sinais na frase que podem ser negados e gera novas sentenças realizando essa substituição. Essa técnica é essencial para aprimorar a capacidade do sistema de tradução em compreender e gerar corretamente sentenças negativas, um aspecto importante da gramática da LIBRAS.
- **Intensidade:** foca na identificação de advérbios na frase que podem ser substituídos para gerar novas frases com diferentes níveis de intensidade. Por exemplo, a substituição de "muito" por "pouco" ou vice-versa.
- **Famosos:** visa diversificar o conteúdo das sentenças ao identificar sinais referentes a pessoas famosas e substituí-los por outros sinais representando outros famosos. Essa técnica melhora a capacidade de desambiguação do sistema ao lidar com pessoas conhecidas.
- **Direcionalidade:** visa capturar uma nuance gramatical específica da LS, levando em consideração o emissor e receptor da ação do verbo, assemelhando-se a concordância número-pessoal. O método de Direcionalidade identifica esses sinais na sentença e realiza substituições apropriadas com outros verbos direcionais equivalentes.

Todas essas técnicas de aumento de dados são aplicadas sobre as sentenças do corpus autêntico construído por linguistas especializados em LIBRAS. As sentenças sintéticas geradas por essas técnicas são então anexadas ao corpus autêntico, formando o corpus padrão utilizado no treinamento.

## 4.3 Treinamento

Os pares de sentenças do corpus são processados por cada método de aumento de dados já implementado no *pipeline* do VLibras, conforme descrito na seção 4.2. O corpus aumentado gerado pela saída do *pipeline* do VLibras compõe o conjunto de dados de treinamento que é o padrão (*baseline*) para os experimentos realizados, possuindo cerca de 106 mil pares de sentenças.

Em seguida, cada método de aumento de dados proposto e implementado neste trabalho é aplicado sobre o conjunto de sentenças do conjunto padrão. Nesta etapa, quando há combinação de diferentes métodos, a entrada de cada método de aumento de dados é restrita às sentenças do conjunto padrão, exceto para o método de retrotradução, devido a restrições de recursos computacionais. No caso da retrotradução, as sentenças sintéticas geradas são anexadas às sentenças autênticas antes da execução dos métodos de aumento de dados padrão do *pipeline* a fim de otimizar recursos e diminuir o custo com o uso de GPU.

A semente de geração de números aleatórios é fixada em todos os experimentos realizados durante o treinamento do modelo, permitindo a reprodutibilidade dos resultados e uma comparação justa entre diferentes treinamentos.

Por fim, o desempenho do modelo é avaliado através de um conjunto de avaliação cuidadosamente selecionado por linguistas, contendo 50 sentenças para cada grupo relevante no domínio da LIBRAS. Esses grupos incluem sentenças básicas, com números cardinais, com palavras homônimas, relacionadas à direcionalidade, a pessoas famosas, à intensidade, a lugares e à negação, permitindo uma avaliação do desempenho do modelo em diferentes contextos linguísticos.

## 4.4 Métricas de Interesse

Avaliar a qualidade das traduções geradas por modelos de NMT é uma tarefa desafiadora devido à natureza complexa e subjetiva da linguagem. Diferentes traduções podem ser consideradas aceitáveis para uma mesma sentença de origem, dependendo de uma variedade de fatores como contexto, estilo e preferências individuais.

Uma das métricas mais amplamente utilizadas para avaliar a qualidade da tradução automática é o BLEU (*Bilingual Evaluation Understudy*) [13]. O BLEU é uma métrica que varia de 0 a 100 e busca automatizar e replicar como um humano julgaria a qualidade da tradução. Essa métrica é baseada na comparação dos n-gramas da sentença gerada com as traduções de referência disponíveis, a fim de calcular a similaridade entre a tradução gerada pelo modelo e a tradução de referência.

O BLEU4, por exemplo, é calculado com base em n-gramas de quatro palavras: Isso significa que o modelo é avaliado com base na precisão dos n-gramas de quatro palavras em suas traduções, em comparação com as traduções de referência. Uma das principais vantagens do BLEU é sua rapidez de cálculo, o que o torna uma métrica eficiente para avaliação automática. No entanto, é importante ressaltar que o BLEU apresenta algumas limitações. Por exemplo, não são consideradas as similaridades semânticas entre as traduções, o que pode levar a pontuações imprecisas em alguns casos.

Neste trabalho, a análise dos resultados é realizada com base no desempenho do modelo de tradução seguindo a métrica BLEU4.

# 5 RESULTADOS E DISCUSSÕES

Foram realizadas duas rodadas de experimentos com o objetivo de avaliar o impacto de diferentes métodos de aumento de dados no desempenho do modelo de tradução. Na primeira rodada, cada método foi aplicado em sequência sobre o conjunto padrão, sem limitar a quantidade de sentenças sintéticas geradas. Na segunda rodada, o tamanho do conjunto de treinamento foi restrito a 155 mil sentenças a fim de equilibrar a proporção entre sentenças autênticas e sintéticas. A Tabela 2 mostra o desempenho do modelo utilizando exclusivamente o corpus autêntico com 65 mil pares de sentenças, sem a aplicação de estratégias de aumento de dados. Em contraste com o corpus padrão, que inclui os métodos de aumento de dados do *pipeline* de tradução do VLibras mencionados na seção 4.2, totalizando 106 mil pares de sentenças.

Todos os métodos de aumento de dados apresentaram melhorias em relação ao conjunto padrão, conforme observado na Tabela 3. A retrotradução trouxe um ganho geral médio significativo (+15,82 BLEU), mesmo com o menor acréscimo de dados (totalizando 155 mil sentenças). Tanto a reversão com substituição de palavras alinhadas, quanto a substituição por sinônimos baseada em *part-of-speech*

**Tabela 2: Resultado em BLEU do desempenho do modelo utilizando apenas o corpus autêntico, sem aumento de dados.**

| Categoria | Pontuação BLEU |
|---|---|
| Básicas | 17,35 |
| Cardinais | 6,22 |
| Contexto | 4,71 |
| Direcionalidade | 0,00 |
| Famosos | 0,00 |
| Intensidade | 0,00 |
| Lugares | 10,79 |
| Negação | 0,00 |
| Romanos | 3,92 |

*tagging* demonstraram resultados próximos (50,73 e 50,32 pontos de BLEU, respectivamente). Observa-se que o aumento de dados traz consideráveis melhorias, analizando sob a perspectiva da métrica BLEU, para todos os grupos de validação.

**Tabela 3: Resultado em BLEU do desempenho do modelo de tradução utilizando as técnicas de aumento de dados selecionadas.**

| | Padrão (106 mil) | RT (155 mil) | RV + SA (313 mil) | SS (346 mil) |
|---|---|---|---|---|
| **Básicas** | 42,05 | 51,66 | 47,89 | 52,67 |
| **Cardinais** | 50,18 | 64,78 | 72,86 | 70,52 |
| **Contexto** | 24,37 | 43,92 | 47,67 | 38,86 |
| **Direcionais** | 0,00 | 18,57 | 30,28 | 29,06 |
| **Famosos** | 27,34 | 45,07 | 43,23 | 40,70 |
| **Intensidade** | 38,47 | 49,38 | 42,5 | 36,55 |
| **Lugares** | 44,04 | 56,68 | 51,85 | 55,04 |
| **Negação** | 44,04 | 50,84 | 57,14 | 62,42 |
| **Romanos** | 25,86 | 57,85 | 63,15 | 67,12 |
| **Média** | 32,93 | 48,75 | **50,73** | 50,32 |

**Legenda: RT (Retrotradução); RV (Reversão); SA (Substituição alinhada); SS (Substituição por sinônimos)**

Ao adicionar a retrotradução em conjunto com a combinação de reversão com substituição alinhada, conforme a Tabela 4, foi observada uma pequena piora não significativa no resultado geral. Por outro lado, a combinação de retrotradução com a substituição por sinônimos resultou em um ganho expressivo de desempenho (+19,25 BLEU em relação ao conjunto padrão), possivelmente devido ao aumento substancial na quantidade de sentenças resultante dessa combinação (totalizando 540 mil sentenças). No entanto, ao incluir mais métodos de aumento de dados, mesmo que isso aumente ainda mais a quantidade de sentenças no conjunto de treinamento, não foi observada uma melhoria significativa no desempenho do modelo. Isso sugere que há uma limitação na melhoria do desempenho proporcionada pela quantidade de sentenças sintéticas geradas, especialmente em função da quantidade de dados autênticos disponíveis. Esses resultados evidenciam a importância de encontrar

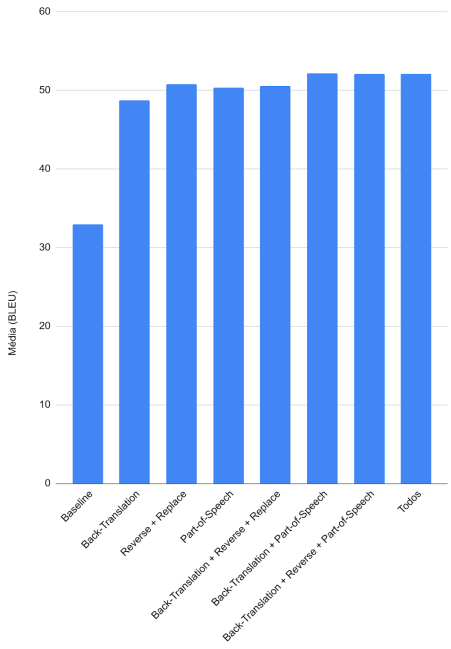


**Figura 3: Média em BLEU do desempenho das técnicas de aumento de dados sem restrição no tamanho do conjunto de treinamento.**

um equilíbrio adequado entre a quantidade de dados autênticos e sintéticos no conjunto de treinamento.

Na segunda fase dos experimentos, conforme exibido na Tabela 5, onde o tamanho do conjunto de treinamento foi limitado, o método de retrotradução obteve os melhores resultados. Esse resultado não é surpreendente, uma vez que a retrotradução é um dos métodos mais estabelecidos e amplamente utilizados para aumento de dados em PLN.

Ao considerar um cenário com aumento de 25% nos dados para cada método empregado, foram testadas diversas combinações de métodos aos pares, conforme visto na Tabela 6. Notadamente, a combinação de retrotradução e substituição por sinônimos com *part-for-speech tagging* demonstrou o melhor desempenho quando há a limitação do tamanho do conjunto de treinamento, apresentando ganho de 16,5 BLEU. Essa combinação já havia apresentado bons resultados quando não há limitação no tamanho do conjunto de treinamento, com 19,25 pontos de BLEU sobre o padrão (ver Tabela 4), isso mostra que as duas técnicas produzem resultados bons quando usadas em conjunto. Uma possível explicação para o bom desempenho dessa combinação é o aumento da diversidade do vocabulário proporcionado pela retrotradução. Ao traduzir as sentenças de volta para o idioma original, o conjunto de treinamento passa a contar com um vocabulário mais amplo, o que, por sua vez, aumenta as opções de substituição por sinônimos.

Devido à limitação na quantidade de sentenças, a eficácia do método de reversão foi reduzida quando combinado com outras técnicas. Essa é uma desvantagem que o experimento traz para esse método, porque ele é sugerido como uma tarefa destinada a

**Tabela 4: Desempenho em BLEU de combinações de técnicas de aumento de dados sem restrição na quantidade de sentenças.**

| | Padrão (106 mil) | RT + RV + SA (467 mil) | RT + SS (540 mil) | RT + RV + SS (695 mil) | Todos (851 mil) |
|---|---|---|---|---|---|
| **Básicas** | 42,05 | 48,55 | 48,14 | 50,76 | 53,53 |
| **Cardinais** | 50,18 | 61,54 | 67,29 | 66,79 | 66,49 |
| **Contexto** | 24,37 | 45,68 | 43,73 | 39,36 | 49,25 |
| **Direcionalidade** | 0,00 | 27,66 | 33,60 | 36,95 | 32,66 |
| **Famosos** | 27,34 | 43,74 | 48,09 | 44,83 | 42,85 |
| **Intensidade** | 38,47 | 45,37 | 39,61 | 42,29 | 42,29 |
| **Lugares** | 44,04 | 54,45 | 48,10 | 52,42 | 47,12 |
| **Negação** | 44,04 | 53,84 | 65,94 | 59,17 | 59,34 |
| **Romanos** | 25,86 | 74,17 | 75,19 | 76,07 | 75,21 |
| **Média** | 32,93 | 50,55 | **52,18** | 52,07 | 52,08 |

**Legenda: RT (Retrotradução); RV (Reversão); SA (Substituição alinhada); SS (Substituição por sinônimos)**

**Tabela 5: Resultado em BLEU de cada técnica utilizada isoladamente, com restrição de 155 mil sentenças no conjunto de treinamento.**

| | Padrão | RT | SS | RV | SA |
|---|---|---|---|---|---|
| **Básicas** | 42,05 | 51,66 | 48,74 | 43,07 | 48,51 |
| **Cardinais** | 50,18 | 64,78 | 56,29 | 60,33 | 60,69 |
| **Contexto** | 24,37 | 43,92 | 35,00 | 32,80 | 37,85 |
| **Direcionalidade** | 0,00 | 18,57 | 0 | 0 | 22,34 |
| **Famosos** | 27,34 | 45,07 | 47,41 | 34,36 | 45,23 |
| **Intensidade** | 38,47 | 49,38 | 38,54 | 38,65 | 43,14 |
| **Lugares** | 44,04 | 56,68 | 51,5 | 47,87 | 47,73 |
| **Negação** | 44,04 | 50,84 | 58,87 | 42,95 | 53,6 |
| **Romanos** | 25,86 | 57,85 | 40,72 | 31,07 | 57,25 |
| **Média** | 32,93 | **48,75** | 41,89 | 36,78 | 46,26 |

**Legenda: RT (Retrotradução); RV (Reversão); SA(Substituição alinhada); SS (Substituição por sinônimos)**

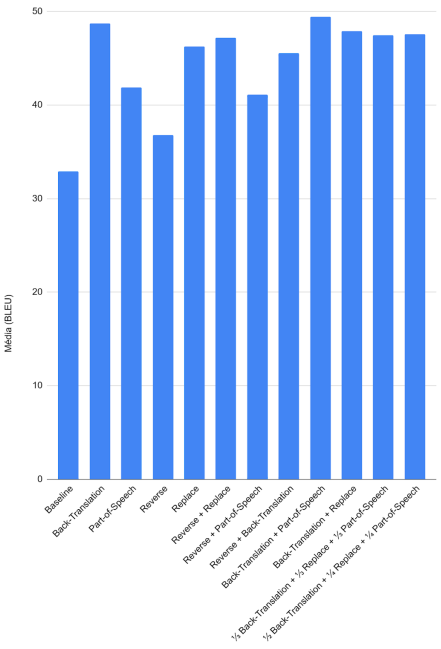


**Figura 4: Média em BLEU do desempenho das técnicas de aumento de dados com limitação de 155 mil sentenças no conjunto de treinamento.**

fortalecer o *encoder* de forma independente de outras técnicas de aumento de dados. Além das combinações de pares de métodos, também foram exploradas outras configurações envolvendo três técnicas distintas.

Por fim, foram testadas mais algumas configurações, conforme mostrado na Tabela 7. Uma combinação de 25% de retrotradução, 25% de substituição por sinônimos e 25% de substituição alinhada e uma configuração com 1/3 de contribuição por cada método. Observa-se, no entanto, que essas combinações não resultaram em melhorias significativas em relação aos resultados obtidos anteriormente. Sendo assim, mesmo com a combinação de diversas técnicas de aumento de dados, a quantidade de sentenças ainda é um fator limitante para alcançar melhorias de desempenho.

**Tabela 6: Resultado em BLEU do treinamento do modelo de tradução com contribuição de 25 mil sentenças por método.**

| | Padrão | RV + SA | RV + SS | RT + SS | RT + RV | RT + SA |
|---|---|---|---|---|---|---|
| **Básicas** | 42,05 | 47,29 | 40,84 | 48,37 | 47,86 | 46,25 |
| **Cardinais** | 50,18 | 66,69 | 57,17 | 66,33 | 67,18 | 64,80 |
| **Contexto** | 24,37 | 43,86 | 33,08 | 38,86 | 38,56 | 43,28 |
| **Direcionalidade** | 0,00 | 18,61 | 17,12 | 28,12 | 21,91 | 25,01 |
| **Famosos** | 27,34 | 52,12 | 41,58 | 46,67 | 42,07 | 46,46 |
| **Intensidade** | 38,47 | 40,21 | 49,63 | 43,62 | 44,40 | 42,06 |
| **Lugares** | 44,04 | 48,65 | 47,97 | 50,84 | 52,17 | 52,65 |
| **Negação** | 44,04 | 57,69 | 48,55 | 60,05 | 46,69 | 50,59 |
| **Romanos** | 25,86 | 49,51 | 34,07 | 62,09 | 48,99 | 60,08 |
| **Média** | 32,93 | 47,18 | 41,11 | **49,43** | 45,53 | 47,90 |

**Legenda: RT (Retrotradução); RV (Reversão); SA (Substituição alinhada); SS (Substituição por sinônimos)**

**Tabela 7: Resultado em BLEU para combinações envolvendo retrotradução, substituição alinhada e substituição por sinônimos.**

| | Padrão | $\frac{1}{3}$ RT + $\frac{1}{3}$ SA + $\frac{1}{3}$ SS | $\frac{1}{2}$ RT + $\frac{1}{4}$ SA + $\frac{1}{4}$ SS |
|---|---|---|---|
| **Básicas** | 42,05 | 54,24 | 52,80 |
| **Cardinais** | 50,18 | 63,44 | 63,78 |
| **Contexto** | 24,37 | 43,20 | 44,02 |
| **Direcionalidade** | 0,00 | 22,35 | 16,92 |
| **Famosos** | 27,34 | 49,94 | 50,71 |
| **Intensidade** | 38,47 | 42,44 | 33,28 |
| **Lugares** | 44,04 | 51,37 | 51,24 |
| **Negação** | 44,04 | 39,94 | 52,71 |
| **Romanos** | 25,86 | 60,28 | 62,68 |
| **Média** | 32,93 | 47,46 | **47,57** |

**Legenda: RT (Retrotradução); SA (Substituição alinhada); SS (Substituição por sinônimos)**

## 6 CONCLUSÃO

A geração de dados sintéticos para aprimorar modelos de NMT em cenários de poucos recursos (*low-resource language*), especialmente para LS como a LIBRAS, é de suma importância para a acessibilidade e inclusão de pessoas surdas. Dessa forma, este trabalho explorou diferentes métodos de aumento de dados a fim de identificar quais abordagens podem melhorar o desempenho do tradutor, com base em testes computacionais amparados pela métrica BLEU.

Observou-se que a técnica amplamente utilizada de retrotradução também é eficaz na tradução de PB para LIBRAS, trazendo um ganho de 15,82 pontos de BLEU em relação ao conjunto padrão. A combinação de retrotradução e substituição por sinônimos com *part-of-speech tagging* trouxe os melhores resultados em ambos os cenários: sem restrição no tamanho do conjunto de treinamento (+19,25 BLEU sobre o padrão) e também quando o conjunto de treinamento foi limitado a 155 mil sentenças (+16,5 BLEU sobre o padrão). Essas técnicas podem ser utilizadas para aumentar a quantidade de exemplos em conjuntos de sentenças que estejam sub-representados no corpus. Destaca-se também, diante dos resultados alcançados, a importância de manter um equilíbrio entre a quantidade de dados sintéticos gerados em relação à quantidade de dados autênticos no corpus original.

Por fim, enxerga-se que a investigação de técnicas mais custosas, porém potencialmente mais eficazes, pode abrir novas possibilidades para aprimorar ainda mais a qualidade da tradução automática. Modelos de linguagem podem produzir texto fluente em uma variedade de contextos linguísticos, portanto, utilizar esses modelos para gerar dados sintéticos é uma sugestão para trabalhos futuros.

## AGRADECIMENTOS

Os autores agradecem à Secretaria Nacional dos Direitos da Pessoa com Deficiência do Ministério dos Direitos Humanos e da Cidadania pelo apoio financeiro para a realização desta pesquisa.

## REFERÊNCIAS

[1] T. M. U. Araújo. 2012. *Uma solução para geração automática de trilhas em língua brasileira de sinais em conteúdos multimídia.* Tese (Doutorado em Automação e Sistemas). Universidade Federal do Rio Grande do Norte, Natal. https://repositorio.ufrn.br/handle/123456789/15190

[2] Renan Costa e Diego Ramon Silva e Samuel Moreira e Daniel Faustino Souza e Rostand Edson Costa e Tiago Maritan Araújo. 2024. Avaliação do uso de modelos de aprendizagem profunda na tradução automática de línguas de sinais. *Revista Principia - Divulgação Científica e Tecnológica do IFPB* 0, 0 (2024). https://periodicos.ifpb.edu.br/index.php/principia/article/view/8053

[3] Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data Augmentation for Low-Resource Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Regina Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, Vancouver, Canada, 567–573. https://doi.org/10.18653/v1/P17-2090

[4] Steven Y. Feng, Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. 2021. A Survey of Data Augmentation Approaches for NLP. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 968–988. https://doi.org/10.18653/v1/2021.findings-acl.84

[5] Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O. K. Li. 2018. Universal Neural Machine Translation for Extremely Low Resource Languages. *CoRR* abs/1802.05368 (2018). arXiv:1802.05368 http://arxiv.org/abs/1802.05368

[6] Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jessica Rodrigues, and Sandra Aluisio. 2017. Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks. In *Proceedings of the XI Brazilian Symposium in Information and Human Language Technology and Collocated Events (STIL 2017).* Uberlandia, Minas Gerais, Brazil.

[7] Jin Yea Jang, Han-Mu Park, Saim Shin, Suna Shin, Byungcheon Yoon, and Gahgene Gweon. 2022. Automatic Gloss-level Data Augmentation for Sign Language Translation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 6808–6813. https://aclanthology.org/2022.lrec-1.734

[8] Z. Liang, H. Li, and J. Chai. 2023. Sign Language Translation: A Survey of Approaches and Techniques. *Electronics* 12, 12 (2023). https://doi.org/10.3390/electronics12122678

[9] Manuella Aschoff C. B. Lima, Tiago Maritan U. de Araújo, Rostand E. O. Costa, and Erickson S. Oliveira. 2022. A machine translation mechanism of Brazilian Portuguese to Libras with syntactic-semantic adequacy. *Natural Language Engineering* 28, 3 (2022), 271–294. https://doi.org/10.1017/S1351324920000662

[10] Alexandre Magueresse, Vincent Carles, and Evan Heetderks. 2020. Low-resource Languages: A Review of Past Work and Future Challenges. *CoRR* abs/2006.07264 (2020). arXiv:2006.07264 https://arxiv.org/abs/2006.07264

[11] Mieradilijiang Maimaiti, Yang Liu, Huanbo Luan, Zegao Pan, and Maosong Sun. 2021. Improving Data Augmentation for Low-Resource NMT Guided by POS-Tagging and Paraphrase Embedding. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20, 6, Article 107 (aug 2021), 21 pages. https://doi.org/10.1145/3464427

[12] Amit Moryossef, Kayo Yin, Graham Neubig, and Yoav Goldberg. 2021. Data Augmentation for Sign Language Gloss Translation. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, Dimitar Shterionov (Ed.). Association for Machine Translation in the Americas, Virtual. https://aclanthology.org/2021.mtsummit-at4ssl.1

[13] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Pierre Isabelle, Eugene Charniak, and Dekang Lin (Eds.). Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 311–318. https://doi.org/10.3115/1073083.1073135

[14] Alberto Poncelas, Dimitar Shterionov, Andy Way, Gideon Maillette de Buy Wenniger, and Peyman Passban. 2018. Investigating Backtranslation in Neural Machine Translation. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Miquel Esplà-Gomis, Maja Popović, Celia Rico, André Martins, Joachim Van den Bogaert, and Mikel L. Forcada (Eds.). Alicante, Spain, 269–278. https://aclanthology.org/2018.eamt-main.25

[15] Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural Machine Translation for Low-resource Languages: A Survey. *ACM Comput. Surv.* 55, 11, Article 229 (feb 2023), 37 pages. https://doi.org/10.1145/3567592

[16] Víctor M. Sánchez-Cartagena, Miquel Esplà-Gomis, Juan Antonio Pérez-Ortiz, and Felipe Sánchez-Martínez. 2021. Rethinking Data Augmentation for Low-Resource Neural Machine Translation: A Multi-Task Learning Approach. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 8502–8516. https://doi.org/10.18653/v1/2021.emnlp-main.669

[17] Vinícius Veríssimo, Cecília Silva, Vitor Hanael, Caio Moraes, Rostand Costa, Tiago Maritan, Manuella Aschoff, and Thaís Gaudêncio. 2019. A study on the use of sequence-to-sequence neural networks for automatic translation of brazilian portuguese to libras. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*. 101–108.

[18] Jing Wang and Lina Yang. 2022. Effective Data Augmentation Methods for CCMT 2022. In *Machine Translation*, Tong Xiao and Juan Pino (Eds.). Springer Nature Singapore, Singapore, 135–142.

[19] Xinyi Wang, Hieu Pham, Zihang Dai, and Graham Neubig. 2018. SwitchOut: an Efficient Data Augmentation Algorithm for Neural Machine Translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, Brussels, Belgium, 856–861. https://doi.org/10.18653/v1/D18-1100