revista principia

ISSN 2447-9187

https://doi.org/10.18265/2447-9187a2022id8053

ARTIGO ORIGINAL

SUBMETIDO 05/10/2023

APROVADO 15/12/2023

PUBLICADO ON-LINE 12/01/2024

VERSÃO FINAL DIAGRAMADA 30/05/2025

EDITOR ASSOCIADO
Prof. Dr. Francisco Petrônio Alencar de Medeiros

Avaliação do uso de modelos de aprendizagem profunda na tradução automática de línguas de sinais

Renan Paiva Oliveira Costa [1] *

Diego Ramon Bezerra da Silva [2]

Samuel de Moura Moreira [3]

Daniel Faustino Lacerda de Souza [4]

Rostand Edson Oliveira Costa [5]

Tiago Maritan Ugulino de Araujo [6]

[1] renan.paiva@lavid.ufpb.br

[2] diego.silva@lavid.ufpb.br

[3] samuel.moura@lavid.ufpb.br

[4] daniel@lavid.ufpb.br

[5] rostand@lavid.ufpb.br

[6] tiagomaritan@lavid.ufpb.br

Centro de Informática, Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil

* *Autor para correspondência.*

RESUMO: Os modelos recentes de Tradução Automática Neural (*Neural Machine Translation* – NMT) podem ser aplicados a idiomas e domínios de poucos recursos sem limitações significativas. Vários estudos investigam se novas técnicas de NMT também podem ser generalizadas para diferentes contextos, considerando a disponibilidade de dados e infraestrutura computacional. Nesse contexto, o objetivo principal deste estudo foi explorar métodos modernos de NMT e analisar a sua potencial aplicabilidade em cenários de poucos recursos, como os das línguas de sinais. Para uma melhor avaliação, foram adaptados e utilizados alguns modelos promissores identificados no componente de tradução automática da Suíte VLibras e os resultados foram comparados com aqueles fornecidos pela arquitetura LightConv atual, utilizando-se o mesmo *corpus* de treinamento e validação bilíngue Português-Libras, um dos maiores desse tipo disponíveis no mundo, constituído por mais de 70.000 frases geradas por linguistas. Os resultados indicam que a adoção de uma das duas arquiteturas de melhor desempenho – a *Basic Transformer* ou a *ByT5* – ajudaria a melhorar a precisão e a qualidade da tradução da Suíte VLibras, com um aumento percentual de até 12,73% considerando a métrica BLEU.

Palavras-chave: acessibilidade; língua de sinais; línguas de poucos recursos; tradução automática neural; transformers.

Evaluation of the use of deep learning models in sign language machine translation

ABSTRACT: *Recent Neural Machine Translation (NMT) models have shown applicability to low-resource languages and domains without significant limitations. Several studies investigate whether new NMT techniques can be generalized across different contexts, considering data availability and computational resources. In this context, the primary objective of this study was to explore modern NMT methods and analyze their potential applicability*

ISSN 2447-9187

in low-resource scenarios, such as sign languages. For a more thorough evaluation, we adapted and employed some promising models identified in the machine translation component of the VLibras Suite, and the results were compared with those provided by the current LightConv architecture, using the same Portuguese-Libras bilingual training and validation corpus, consisting of over 70,000 sentences generated by linguists, one of the largest of its kind globally. The results indicate that adopting one of the best-performing architectures (Basic Transformer or ByT5) could improve translation accuracy and quality in the VLibras Suite, with a percentage increase of up to 12.73% in the BLEU metric.

Keywords: *accessibility; low-resources languages; neural machine translation; sign language; transformers.*

1 Introdução

A comunidade surda, a qual representa uma parcela relevante da população brasileira e mundial, enfrenta diversos desafios no acesso à informação, normalmente disponibilizada por meio da língua escrita ou falada. Esses desafios se devem, principalmente, ao fato de que a maioria dos surdos passa vários anos na escola, mas não consegue atingir proficiência na leitura e na escrita da língua oral de seu país (Souza *et al*., 2017).

O principal motivo dessa dificuldade é que os surdos se comunicam, naturalmente, por meio de línguas de sinais (LS), sendo as línguas orais (LO) apenas uma segunda língua. Cada LS, por sua vez, é uma língua natural, com léxico e gramática próprios, desenvolvida por cada comunidade de surdos ao longo do tempo, assim como cada comunidade de ouvintes desenvolveu sua língua oral. Essa característica própria de formação da língua faz com que não exista uma língua de sinais única, praticada em todo o mundo. Embora haja muitas similaridades entre essas línguas, cada país normalmente possui a sua própria, e alguns até mais de uma (Quadros, 2006).

Para se garantir um acesso adequado à comunicação, à informação, o ideal é que os conteúdos em línguas orais sejam traduzidos ou interpretados para a LS associada. Considerando-se, contudo, o volume e o dinamismo de informações em alguns ambientes e plataformas, como, por exemplo, na web, fazer isso apenas com o auxílio de intérpretes humanos é tarefa quase inviável, mesmo ao se considerar apenas o conteúdo que é adicionado diariamente na internet. Para abordar essa questão de forma pragmática, uma das estratégias mais promissoras atualmente é a utilização de ferramentas para tradução automática (*machine translation*) de uma língua oral para uma língua de sinais (Corrêa; Cruz, 2019).

Um dos principais desafios dos sistemas de tradução automática para língua de sinais é garantir que o conteúdo disponibilizado aos surdos chegue com a mesma consistência e qualidade do original, permitindo, assim, o entendimento adequado da mensagem (Farooq *et al*., 2021). Tais sistemas são geralmente divididos em quatro classes principais: Tradução Automática Baseada em Regras (*Rule-Based Machine Translation* – RBMT), Tradução Automática Estatística (*Statistical Machine Translation* – SMT), Tradução Automática Baseada em Exemplos (*Example-Based Machine Translation* – EBMT) e Tradução Automática Neural (*Neural Machine Translation* – NMT) (Rivera-Trigueros; Olvera-Lobo; Gutiérrez-Artacho, 2021).

ISSN 2447-9187

Para construir soluções para Processamento de Linguagem Natural (ou NLP, do inglês *Natural Language Processing*) para qualquer idioma, um dos requisitos mais importantes é dispor de dados nesse idioma (Koehn; Knowles, 2017; Ranathunga *et al*., 2023). A Tradução Automática Neural, por exemplo, geralmente baseada em Aprendizagem Profunda (ou DL, do inglês *Deep Learning*), utiliza bases de dados com exemplos de sentenças tanto na língua de origem quanto na língua de destino para o modelo neural aprender a realizar as traduções.

Existem mais de 7.000 idiomas falados em todo o mundo, mas, desse total, apenas cerca de 20 possuem corpo (ou *corpus*[1]) de centenas de milhões de palavras (Dryer; Haspelmath, 2013). O inglês é um dos idiomas com maior quantidade de dados, seguido do chinês e do espanhol. Outros idiomas com grandes conjuntos de dados incluem as línguas da Europa Ocidental e o japonês (Lewis, 2014). Por outro lado, a maioria dos idiomas falados na Ásia e na África não possuem os dados de treinamento necessários para se construírem sistemas NLP precisos. Essas linguagens são chamadas de linguagens de poucos recursos (do inglês *low resources languages*[2]) (Khan *et al*., 2023). Esse também é o caso da maioria das línguas de sinais, caracterizadas pela quase inexistência de material oralizado natural (escrito ou falado) em LS e, frequentemente, pela escassez de *corpora* bilíngues que, quando existentes, são de pequeno porte, geralmente produzidos por linguistas.

[1] Um *corpus* é um conjunto de dados linguísticos, seja em forma escrita ou oral, que serve como base para a análise da língua. Quando um *corpus* possui um conjunto de sentenças equivalentes em mais de uma língua, ele é denominado de *corpus* bilíngue. Os conteúdos padrão em várias línguas, como, por exemplo, a Bíblia, são uma ótima referência para a construção de *corpora* bilíngues.

[2] Tecnicamente, sempre que uma linguagem carece de grandes *corpora* monolíngues ou bilíngues ou de recursos linguísticos suficientes criados manualmente para a construção de modelos de NLP, ela é considerada uma linguagem de poucos recursos.

Uma parcela significativa da população mundial, incluindo a comunidade surda, ainda é mal atendida pelos sistemas NLP, devido a vários desafios que os desenvolvedores enfrentam ao construir sistemas para linguagens de poucos recursos, como as línguas de sinais (Haque; Liu; Way, 2021; Khan *et al*., 2023). Entre esses desafios, destacam-se:

- Falta de conjuntos de dados anotados: conjuntos de dados anotados são necessários para treinar modelos de aprendizagem profunda (DL) de maneira supervisionada. Esses modelos são comumente usados para resolver, com muita precisão, tarefas específicas, como, por exemplo, a detecção de discurso de ódio. A criação de conjuntos de dados anotados requer, no entanto, intervenção humana, rotulando exemplos de treinamento um por um, o que torna o processo geralmente demorado e oneroso, dados os milhares de exemplos exigidos pelos modelos avançados de aprendizado profundo. Assim, torna-se inviável contar apenas com a criação manual de dados a longo prazo;
- Falta de conjuntos de dados não rotulados: conjuntos de dados não rotulados, como *corpus* de texto, são os precursores de suas versões anotadas. Eles são essenciais para treinar modelos básicos, que posteriormente são ajustados para tarefas específicas. Abordagens para contornar a falta de conjuntos de dados não rotulados tornam-se, portanto, igualmente importantes;
- Carência de suporte a vários dialetos de um idioma: línguas com múltiplos dialetos também representam um problema complicado de resolver, especialmente para modelos de fala. Um modelo treinado em um idioma geralmente não terá um ótimo desempenho em seus diferentes dialetos. Por exemplo, a maioria dos conjuntos de dados não rotulados e anotados disponíveis para o árabe está em árabe padrão moderno, no entanto, em assistentes de voz ou bate-papo para uso diário, esse padrão é considerado muito formal para muitos falantes de árabe. Assim, o suporte a dialetos torna-se necessário para casos de uso prático.

Algumas pesquisas recentes em NLP de poucos recursos (*low-resources* NLP) buscam adaptar soluções de NLP existentes, baseadas em línguas com mais dados, para idiomas e domínios que carecem de recursos. A premissa é que os modelos modernos de NLP

ISSN 2447-9187

podem ser igualmente aplicáveis tanto para linguagens de poucos recursos quanto para línguas com abundância de dados e, possivelmente, podem ser generalizados com sucesso para cenários com diferentes níveis de recursos, tanto em termos de disponibilidade de dados quanto de recursos computacionais.

Nesse contexto, o foco deste trabalho é avaliar alguns métodos em evidência, baseados em *deep learning* e empregados na tradução automática de *low resources languages*, e analisar sua potencial aplicabilidade para a tradução de línguas de sinais. O objetivo principal é identificar quais modelos, entre os considerados, podem se adequar melhor para a tradução de Português Brasileiro para glosas[3] em Libras, a Língua Brasileira de Sinais. Para uma melhor avaliação, alguns dos modelos mais promissores identificados na literatura foram adaptados e utilizados no componente tradutor da Suíte VLibras[4], e os resultados obtidos foram comparados com aqueles fornecidos atualmente pela ferramenta, visando avaliar se as novas abordagens podem representar alternativas para a melhoria na qualidade da tradução Português-Libras disponível atualmente.

O restante deste artigo está organizado da seguinte forma. Na Seção 2, são apresentados os trabalhos relacionados à temática deste estudo. Na Seção 3, são descritos os modelos de DL selecionados para o estudo e os critérios de elegibilidade utilizados. Na Seção 4, faz-se o relato do processo de desenvolvimento, da avaliação dos modelos candidatos e dos principais resultados obtidos. Por fim, as conclusões são apresentadas na Seção 5.

[3] Glosas são palavras de uma determinada língua oral, grafadas em letras maiúsculas, que representam sinais manuais de sentido próximo. Wilcox e Wilcox (1997) definem glosa como uma tradução simplificada de morfemas da língua de sinais para morfemas de uma língua oral.

[4] A Suíte VLibras (Araújo, 2012) é o resultado de uma parceria entre o Ministério de Planejamento, Desenvolvimento e Gestão (MP), intermediada pela Secretaria de Tecnologia da Informação (STI) e pela Universidade Federal da Paraíba (UFPB), representada esta última pelo Laboratório de Aplicações de Vídeo Digital (LAVID). A Suíte VLibras consiste em um conjunto de ferramentas gratuitas e de código aberto para tradução automática de Português Brasileiro (texto, áudio e vídeo) para a Língua Brasileira de Sinais (Libras), tornando computadores, dispositivos móveis e plataformas web acessíveis para os surdos. Atualmente, o VLibras é usado em mais de 500.000 sites públicos e privados, incluindo os principais sites do Governo Brasileiro (brasil.gov.br), da Câmara dos Deputados (camara.leg.br) e do Senado Federal (senado.leg.br), estando presente no cotidiano da comunidade surda com milhões de traduções mensais. Mais informações podem ser obtidas em http://www.vlibras.gov.br.

2 Trabalhos relacionados

Nesta seção, são apresentados alguns trabalhos relacionados ao tema desta pesquisa, além dos modelos utilizados e/ou analisados nesses trabalhos, que foram pré-selecionados como candidatos para a avaliação proposta.

No contexto de modelos baseados exclusivamente em redes neurais profundas, Shazeer *et al*. (2017) propuseram um mecanismo denominado *Mixture of Experts* (MoE), que consiste em um número do que os autores denominaram de *experts*, que se traduzem em um conjunto de redes neurais *feed-forward*, combinadas em uma rede de bloqueio que seleciona combinações esparsas dos ditos *experts* para processar cada entrada. O mecanismo MoE é aplicado intermediariamente em uma pilha de redes LSTM (*Long Short-Term Memory).* Entre os avanços apresentados por Shazeer *et al*. (2017), no contexto de tradução, destacam-se valores de BLEU de 40,56% para tradução inglês-francês (En-Fr), usando-se a base WMT'14/En-Fr, e 26,03% para tradução Inglês-Alemão (En-De) usando-se a base WMT'14/En-De.

No contexto de idiomas com poucos recursos (*low resource languages*), Ortega, Mamani e Cho (2020) propuseram um sistema NMT baseado em LSTM e com um mecanismo de segmentação morfológica baseado em *Byte Pair Encoding* (BPE) (Gage *et al*., 1994).

Na linha de modelos que se utilizam dos mecanismos de atenção, observa-se um número crescente de trabalhos que seguem essa abordagem. Um mecanismo de reconhecimento de sinais em vídeo e posterior tradução para linguagem falada foi proposto por Camgoz *et al*. (2018). Para a etapa de tradução dos *tokens* provenientes do processamento de vídeo, os autores utilizaram uma Rede Neural Recorrente (*Recurrent Neural Network* ou RNN) com mecanismo de atenção para realizar a etapa de tradução de glosa para texto.

Arvanitis, Constantinopoulos e Kosmopoulos (2019) tratam do problema de tradução de glosa para texto partindo da *American Sign Language* (ASL) para o inglês, utilizando três diferentes funções de atenção para construção da solução. No mesmo tema, Amin, Hefny e Mohammed (2021) propuseram uma abordagem bidirecional com base em *Gated Recurrent Units* (GRU), *Long Short-Term Memory* (LSTM) e mecanismo de atenção. Os autores aplicaram o modelo para tradução bidirecional entre a Língua Inglesa e a Língua Americana de Sinais (ASL). Outros trabalhos na mesma temática foram desenvolvidos por Abujar *et al*. (2021), Hamed, Helmy e Mohammed (2022), Yonglan e Wenjia (2022) e Zhang e Duh (2021).

No campo de soluções inteiramente baseadas em mecanismos de atenção, a arquitetura *Transformer* se apresenta como o estado da arte em termos de melhores resultados para o problema de tradução automática. Muitos trabalhos têm sido desenvolvidos à luz dessa arquitetura e de modelos dela derivados.

No contexto de tradução de línguas faladas para línguas de sinais, Camgoz *et al*. (2020) propuseram o uso de *transformers* para resolver o problema de tradução de texto para glosa. Yin e Read (2020) empregaram modelos baseados em *transformers* e *Spatial-Temporal Multi-Cue* (SMTC) para realizar as tarefas de reconhecimento e tradução de sinais. Na mesma direção, uma arquitetura denominada *Progressive Transformers* foi apresentada por Saunders, Camgoz e Bowden (2020), com foco na tradução de texto para sequências contínuas de poses tridimensionais de sinais. O trabalho de Gómez, McGill e Saggion (2021) propôs o uso de *transformers* para o processo de tradução de texto para glosa, com uma etapa de pré-processamento que considera informações de dependência léxica para tal processo. Outros estudos sobre reconhecimento e tradução de sinais para texto e texto para glosa foram desenvolvidos por Angelova, Avramidis e Möller (2022) e Mohamed, Hefny e Amin (2022).

Alguns dos novos modelos *transformers,* ainda pouco explorados ou não testados, se mostram promissores para aplicação no problema de tradução de texto para glosa. Essas arquiteturas serão discutidas com maior profundidade na Seção 3 deste artigo, com destaque para os modelos *Bidirectional Encoder Representations from Transformers* (BERT), *Bidirectional and Auto-Regressive Transformer* (BART) e *Text-to-Text Transfer Transformer* (T5).

3 Seleção de modelos candidatos

Nesta seção, são dispostos os critérios de elegibilidade utilizados na seleção dos modelos candidatos para avaliação, entre os identificados na revisão da literatura.

3.1 Critérios de elegibilidade

Desde a introdução da arquitetura LightConv (modelo atual do tradutor neural do VLibras) em 2019, novas técnicas e modelos têm sido propostos na literatura. A popularidade das arquiteturas baseadas em *transformers* tem aumentado e, atualmente, a maioria dos problemas e tarefas de processamento de linguagem natural apresenta seu estado da arte baseado nessas redes.

A revisão da literatura realizada para a prospecção de modelos, descrita na Seção 2, teve como objetivo identificar quais modelos foram mais aplicados e/ou referenciados

em artigos recentes da área, publicados entre 2017 e 2023 e relacionados ao tema em pauta, principalmente em *low resource* NLP e/ou *sign language* NMT.

Para essa fase de experimentação, foram definidos critérios adicionais de inclusão e exclusão para a seleção dos modelos candidatos a uma avaliação mais detalhada. Assim, além do possível ganho em qualidade na tradução, outros fatores também foram considerados na escolha dos modelos candidatos, incluindo:

- Custo de infraestrutura de treinamento;
- Reprodutibilidade;
- Viabilidade de expansão e customização dos modelos;
- Ausência de restrições para uso e licenciamento.

3.2 Modelos candidatos

Entre as diversas arquiteturas e variações da arquitetura *transformer*, algumas são geralmente consideradas mais promissoras para problemas de tradução automática. Partindo dos modelos mais referenciados nos trabalhos e considerando os critérios já mencionados, foram pré-selecionados e tiveram a viabilidade da experimentação verificada apenas os modelos disponíveis no portal PapersWithCode[5], que reúne um acervo de trabalhos de pesquisa reprodutíveis, oferecendo *datasets*, códigos-fonte e *benchmarks* comparáveis, obtidos sobre *corpora* públicos relevantes.

[5] Disponível em: https://paperswithcode.com/. Acesso em: 27 dez. 2023.

Após a confirmação de viabilidade de experimentação, foram selecionados os seguintes modelos para avaliação criteriosa:

- *Transformer* básico (ou *Vanilla Transformer*);
- BERT, da Google;
- BART, da Meta;
- T5 e ByT5, da Google.

As arquiteturas BART e T5 foram escolhidas devido à disponibilidade de modelos pré-treinados em grandes *corpora*, facilitando tarefas de processamento de linguagem natural, como a tradução automática. Em especial, a T5 possui versões treinadas no *corpus* BrWac (Wagner Filho *et al*., 2018), um grande *corpus* de Português Brasileiro. Embora não existam versões generalistas para essa língua, a arquitetura BART conta com versões treinadas para múltiplos idiomas, incluindo o Português (Liu; Winata; Fung, 2021). A arquitetura ByT5, por sua vez, herda as características da T5, além de apresentar um processo de tokenização mais neutro em relação a especificidades e resiliente a ruídos.

Ademais, os novos modelos também foram avaliados quanto ao custo computacional e de infraestrutura. A partir desses critérios, foram considerados apenas modelos que pudessem ser executados em ambientes (servidores) baseados em CPUs. Esse é possivelmente o principal requisito não funcional do componente de tradução de uma solução como a Suíte VLibras, para tornar viável a sua operação como uma plataforma gratuita e de amplo acesso. O valor de referência para o tempo de processamento de uma tradução na infraestrutura atual do VLibras foi estimado, em média, em 1,2 segundos. No contexto dessa avaliação, um modelo candidato será considerado inviável se seu tempo de inferência em CPU for superior a 2 segundos.

ISSN 2447-9187

4 Avaliação dos modelos candidatos

Esta seção, além de descrever como a avaliação experimental dos modelos selecionados foi planejada e realizada, também faz a apresentação e discussão dos resultados obtidos.

4.1 Planejamento de experimentos

As subseções a seguir detalham o planejamento e a condução da avaliação experimental proposta para o estudo, incluindo a adaptação dos modelos candidatos para utilização na Suíte VLibras.

4.1.1 Metodologia

O objetivo da experimentação é testar os modelos *Transformer* básico, BART, BERT, T5 e ByT5 diretamente nos componentes de tradução do VLibras. Os resultados obtidos são comparados com os gerados pela versão atual do tradutor híbrido do VLibras, que utiliza o modelo LightConv do *framework* Fairseq[6].

[6] Fairseq é um kit de ferramentas de modelagem de sequência, produzido pela Meta, para o treinamento de modelos personalizados voltados para tradução, resumo e outras tarefas de geração de texto.

Para permitir a comparação com os resultados já disponíveis do VLibras, utiliza-se nos experimentos o mesmo *corpus* de treinamento e validação e calculam-se as mesmas métricas de avaliação. Esse *corpus* é o mesmo utilizado no treinamento do *pipeline* de tradução do VLibras atualmente em produção. Esse *dataset* possui mais de 70.000 tuplas de português/glosa, desenvolvidas manualmente por linguistas e intérpretes, sendo um dos maiores *corpora* desse tipo disponíveis no mundo.

Para facilitar a execução dessa etapa de experimentação, foi utilizada a biblioteca de aprendizado profundo para NLP Hugging Face[7], que oferece acesso a vários modelos pré-treinados de processamento de linguagem natural, como modelos de linguagem e tokenizadores. A biblioteca permite realizar tarefas comuns de NLP, como classificação de texto, extração de entidades e tradução automática, entre outras, e oferece ferramentas para treinar e personalizar modelos para tarefas específicas, integrando-se facilmente com outras bibliotecas e *frameworks*. Essa biblioteca foi selecionada devido à disponibilidade dos modelos prospectados, à facilidade de integração e por estar em constante desenvolvimento, garantindo, assim, boa manutenção para o projeto.

[7] Disponível em: https://huggingface.co. Acesso em: 27 dez. 2023.

4.1.2 Arquitetura de tradução da Suíte VLibras

O componente de tradução do VLibras atualmente adota uma arquitetura híbrida baseada em um tradutor de regras (RBMT) (Oliveira *et al*., 2019) e um tradutor baseado em inteligência artificial (NMT), usando o modelo LightConv (Wu *et al*., 2019). Nesse contexto, o tradutor de regras atua como um componente de pré-processamento da sentença em Português, componente que, por sua vez, alimenta o modelo LightConv, visando normalizar a entrada e ajudar o modelo durante o treinamento. Essa etapa é fundamental em função do volume relativamente baixo de dados disponíveis.

O fluxo de inferência do processo de tradução (Figura 1) é representado pelas etapas que uma frase em Português percorre até ser convertida em uma representação traduzida

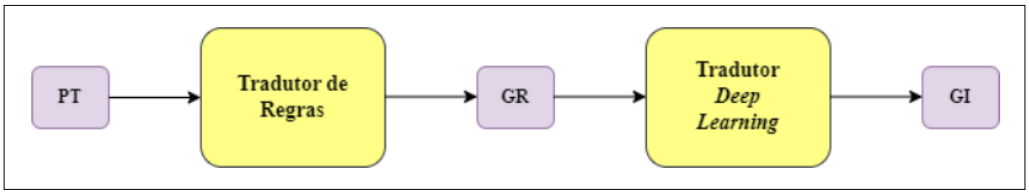
ISSN 2447-9187

em glosa pronta para ser consumida por outra aplicação. As etapas que o VLibras atualmente utiliza para a inferência são:

- Receber a frase em Português (PT);
- Inserir a frase no tradutor baseado em regras;
- Gerar uma Glosa Intermediária[8] (GR);
- Inserir a Glosa Intermediária no tradutor neural;
- Gerar a Glosa Final (GI), pronta para ser sinalizada.

[8] Frase simplificada e em estado intermediário, destinada a facilitar a tradução por *deep learning*.

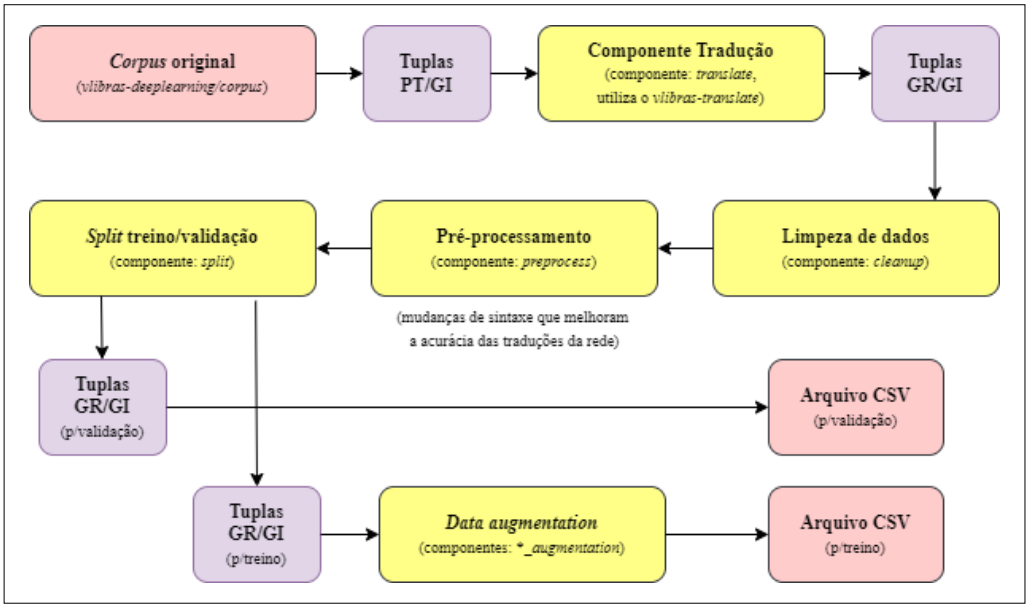
Figura 1 ▶
Fluxo de transformações do tradutor do VLibras.
Fonte: elaborado pelos autores



O treinamento do tradutor *deep learning* do VLibras também segue um *pipeline* (Figura 2). O *corpus* bilíngue de treinamento, um conjunto de dados com diversas sentenças equivalentes em Português para glosa, é traduzido utilizando-se o tradutor de regras para glosa intermediária. Nesse processo, as sentenças passam por várias etapas de pré-processamentos e, em seguida, são divididas em dois conjuntos distintos: um para o treinamento do modelo e outro para a validação do processo de tradução. O conjunto de treinamento também passa por um processo de *data augmentation*[9], para ampliar a ocorrência de palavras raras em seu conjunto de sentenças.

[9] *Data augmentation* é um recurso amplamente utilizado em NLP de poucos recursos (*low-resource* NLP) para ampliar, sinteticamente e por meio de técnicas específicas, a quantidade de sentenças em *corpora* usados no treinamento de modelos neurais.

Figura 2 ▶
Arquitetura geral do fluxo de treinamento do VLibras.
Fonte: elaborado pelos autores



Após o pré-processamento do *corpus*, os dados são submetidos a um componente de aprendizado para geração de *tokens* BPE[10], para serem aplicados nas tuplas de treino e validação. Em seguida, as tuplas são binarizadas para utilização no treinamento do modelo usado no tradutor *deep learning* atual do VLibras (modelo LightConv do *framework* Fairseq).

[10] *Byte pair encoding* (BPE) é um método de tokenização que representa um texto com o menor número de bytes.

4.1.3 Métricas de interesse

Os modelos de *deep learning* são treinados em grandes bases de dados, chamadas *corpora*, e é necessária a avaliação automática dessas traduções para medir a eficiência do modelo de tradução automática. Até o presente momento, a métrica de avaliação mais usada para esse fim é a BLEU, proposta originalmente por Papineni *et al*. (2002). A estratégia da métrica BLEU consiste em calcular a similaridade semântica entre a tradução gerada pelo computador e uma ou mais traduções humanas de referência, sendo projetada para substituir e automatizar a avaliação humana em cenários que exigem múltiplas avaliações.

O resultado da métrica BLEU geralmente é expresso como um número entre 0 e 1, em que 1 indica uma correspondência perfeita entre a tradução gerada e a tradução de referência. Valores mais próximos de 1 indicam melhores resultados. Alguns algoritmos de tradução automática são avaliados com o *corpus* de dados de avaliação BLEU, e essa métrica tem demostrado eficiência em indicar o desempenho dos modelos de tradução automática.

Além da métrica BLEU, também é possível avaliar traduções automáticas utilizando medidas de similaridade. A distância de Levenshtein (Levenshtein, 1966), conhecida também como distância de edição, é uma medida de similaridade entre duas sequências de caracteres (*strings*). Ela se baseia no número mínimo de operações de edição (inserção, deleção ou substituição de caracteres) necessárias para transformar uma *string* em outra.

Uma variação da distância de Levenshtein é a sua versão normalizada, que calcula a distância padrão, mas normaliza o resultado dividindo-o pelo comprimento da *string* mais longa. Dessa forma, a distância de Levenshtein normalizada varia de 0 a 1, em que 0 indica que as sentenças não compartilham nenhuma palavra ou *token* em comum, e 1 indica que as sentenças são idênticas.

A distância de Levenshtein normalizada é amplamente usada como métrica de similaridade para *strings* e se aplica a diversas áreas, como detecção de plágios, processamento de linguagem natural, reconhecimento de fala e tradução automática. Essa normalização é útil, pois permite uma comparação justa entre *strings* de diferentes tamanhos, evitando que a métrica seja excessivamente sensível às diferenças de comprimento.

Para avaliação do VLibras, uma tradução com distância de Levenshtein normalizada igual a 1 é considerado correta, representando uma tradução perfeita. Uma distância de Levenshtein normalizada inferior a 0,85 indica uma tradução incorreta, enquanto valores entre 0,85 e 1 indicam uma tradução parcialmente correta. Esses valores limiares foram estabelecidos empiricamente durante o desenvolvimento do componente de tradução do VLibras, com base em testes, avaliações e na percepção de qualidade da tradução por parte de pessoas surdas, intérpretes e linguistas, tanto da equipe do VLibras quanto usuários externos. Resumidamente:

- Ok (similaridade igual a 1);
- Parcial (similaridade menor que 1 e maior ou igual a 0,85);
- Incorreta (similaridade menor que 0,85).

Assim, a avaliação computacional usada no componente tradutor do VLibras utiliza duas métricas principais: uma métrica de tradução (BLEU) e uma métrica de

similaridade (distância de Levenshtein normalizada). Essas métricas serão adotadas neste estudo para permitir uma comparação consistente.

4.1.4 Conjuntos de avaliação

Tão importante quanto a definição das métricas de interesse é a seleção do conjunto de dados que será utilizado para avaliar o modelo, conhecido como conjunto de avaliação. Esse conjunto é um subconjunto do conjunto de dados de treinamento separado e usado exclusivamente para avaliar o desempenho de um modelo de *deep learning*, medindo sua capacidade de generalizar para dados inéditos.

Geralmente, os dados disponíveis (neste caso em particular, o *corpus* bilíngue de referência) são divididos em três conjuntos: i) treinamento; ii) validação; iii) avaliação. O conjunto de treinamento é usado para treinar o modelo, enquanto o de validação permite selecionar o melhor modelo entre várias configurações (por exemplo, ajustando hiperparâmetros). O conjunto de avaliação serve para medir o desempenho final do modelo selecionado.

É fundamental que o conjunto de avaliação seja totalmente separado dos conjuntos de treinamento e de validação, de forma que esse conjunto contenha dados inéditos para o modelo, evitando que o modelo "memorize" os dados de treinamento e validação, o que resultaria em uma superestimação de seu desempenho, fenômeno conhecido como sobreajuste dos dados de treinamento, ou *overfitting*.

Também é relevante avaliar o modelo em diferentes conjuntos de dados, para testar tanto seu desempenho quanto sua robustez. Para isso, as avaliações devem ser feitas com diferentes tipos de dados, como dados desbalanceados, dados incompletos, dados diferentes daqueles usados no treinamento etc. Isso permite compreender melhor como o modelo se comporta e identificar potenciais limitações.

No caso do VLibras, a avaliação é realizada sobre diferentes conjuntos de teste, que procuram modelar cenários e pontos críticos identificados no processo de tradução de português para Libras, sendo projetados com a supervisão de especialistas em Libras. Esses conjuntos são formados de:

- frases básicas;
- frases contendo referências de contexto;
- frases com referências direcionais;
- frases com negação;
- frases com nomes de pessoas famosas;
- frases com referências de lugares;
- frases com indicadores de intensidade;
- frases com números cardinais;
- frases com números romanos.

Todos os modelos de tradução automática de português para Libras gerados no contexto do VLibras são avaliados nesses conjuntos antes de serem homologados e disponibilizados para o usuário final por meio dos componentes interativos da Suíte VLibras.

4.1.5 Configuração do ambiente

Para realizar os experimentos, foram utilizados dois ambientes de processamento, a fim de permitir a execução paralela de cada etapa planejada, considerando que cada ciclo de treinamento e validação durava aproximadamente cinco horas. A configuração de cada ambiente exigiu a instalação de diversos módulos Python dos *frameworks* utilizados. Em seguida, foi baixado o código-fonte do *pipeline* do VLibras em produção e seus submódulos, hospedados no GitLab do Laboratório de Aplicações de Vídeo Digital da UFPB (LAVID). Os dois ambientes foram configurados de maneira similar para o treinamento de cada modelo previsto, incluindo a adaptação e a integração do modelo ao *pipeline* do VLibras.

Antes da execução dos experimentos, os hiperparâmetros foram verificados para garantir conformidade com a versão de produção, e testes de sanidade foram conduzidos para aferir se ambos os ambientes forneciam resultados compatíveis e sincronizados. Adicionalmente, foram executados treinamentos exploratórios e comparados aos resultados do modelo atual, com ajustes necessários, como atualização de dependências e variáveis de ambiente, entre outros, até que os resultados fossem equivalentes.

Durante essa fase, foi identificado que o modelo BERT não apresentava resultados adequados no contexto de tradução Português-Libras. De maneira geral, a reprodutibilidade dos resultados relatados na literatura era limitada, especialmente no caso da tradução automática, o que levou ao seu descarte nas etapas seguintes de avaliação.

4.1.6 Realização dos experimentos

O treinamento e a avaliação de cada experimento foram executados em paralelo em um dos dois ambientes, e os resultados foram calculados e consolidados para cada modelo e cada subconjunto de sentenças de avaliação. O objetivo do primeiro ciclo de experimentos foi estabelecer uma pontuação de referência do modelo atual, LightConv. Nos ciclos seguintes, foram testados os modelos BART, *Transformer* básico, T5 e ByT5, aos quais algumas técnicas também foram combinadas, como *back translation*, aplicação de *data augmentation* antes da tradução para glosa intermediária no pré-processamento e alteração na quantidade de *tokens* BPE.

4.2 Análise de resultados

Nesta subseção, são apresentados os resultados obtidos em cada ciclo de experimentos. As tabelas de resultados, em geral, possuem uma coluna para um dos nove subconjuntos de avaliação considerados e uma linha para cada categoria de classificação da tradução (OK, Parcial e Incorreto). Os valores em cada célula indicam o percentual de resultados de cada classificação alcançado em cada subconjunto pelo modelo/configuração em análise.

O modelo atualmente em uso no VLibras, um tradutor híbrido baseado no modelo LightConv, é utilizado como referência para a avaliação de novos modelos. Dessa forma, é possível verificar se os novos modelos apresentam métricas computacionais superiores ou inferiores ao modelo em produção (Figura 3).

ISSN 2447-9187

Figura 3 ▶
Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo de referência LightConv.
Fonte: dados da pesquisa

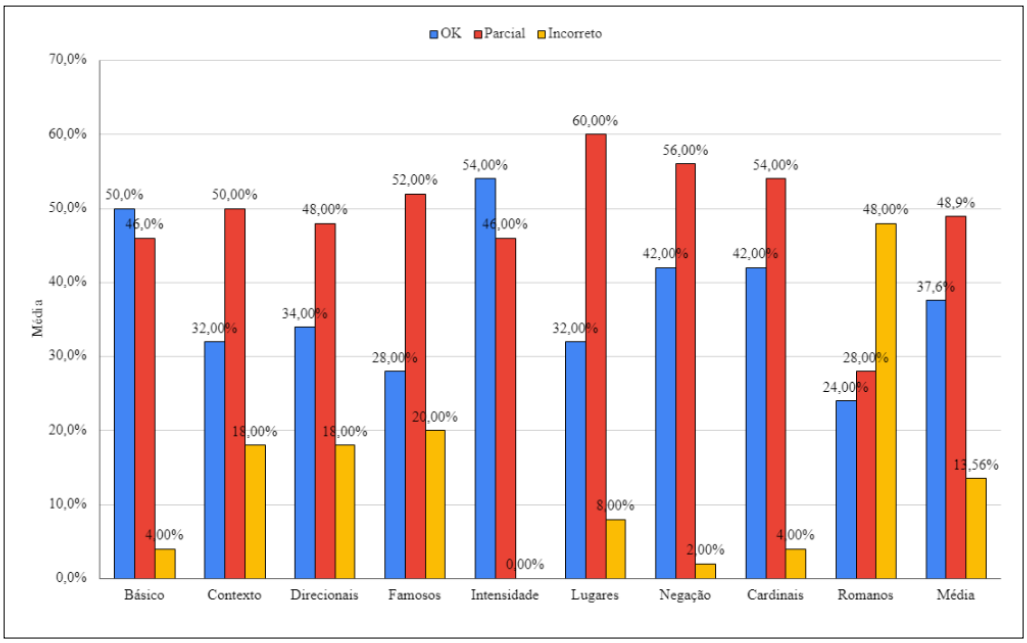
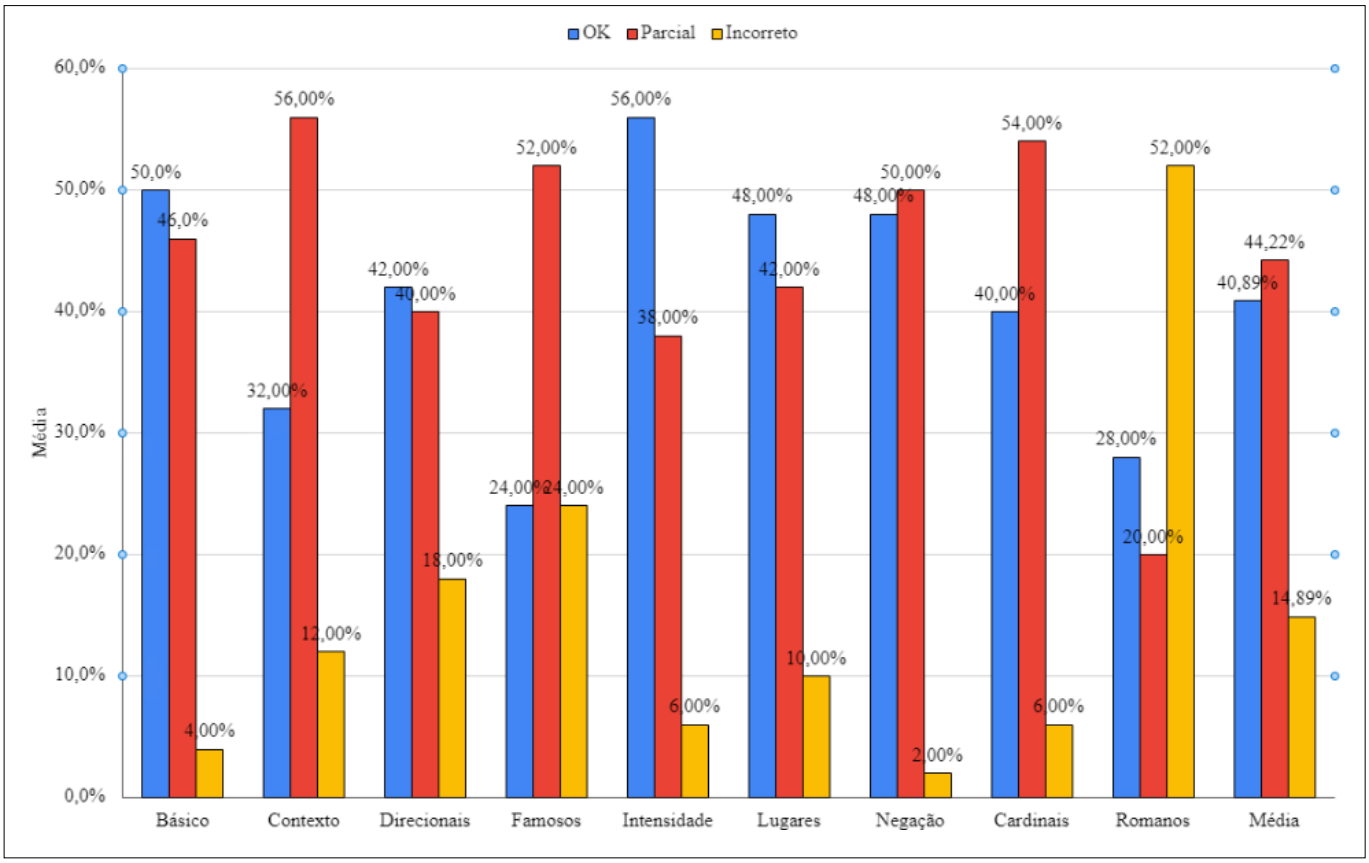


Figura 4 ▼
Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo BART.
Fonte: dados da pesquisa

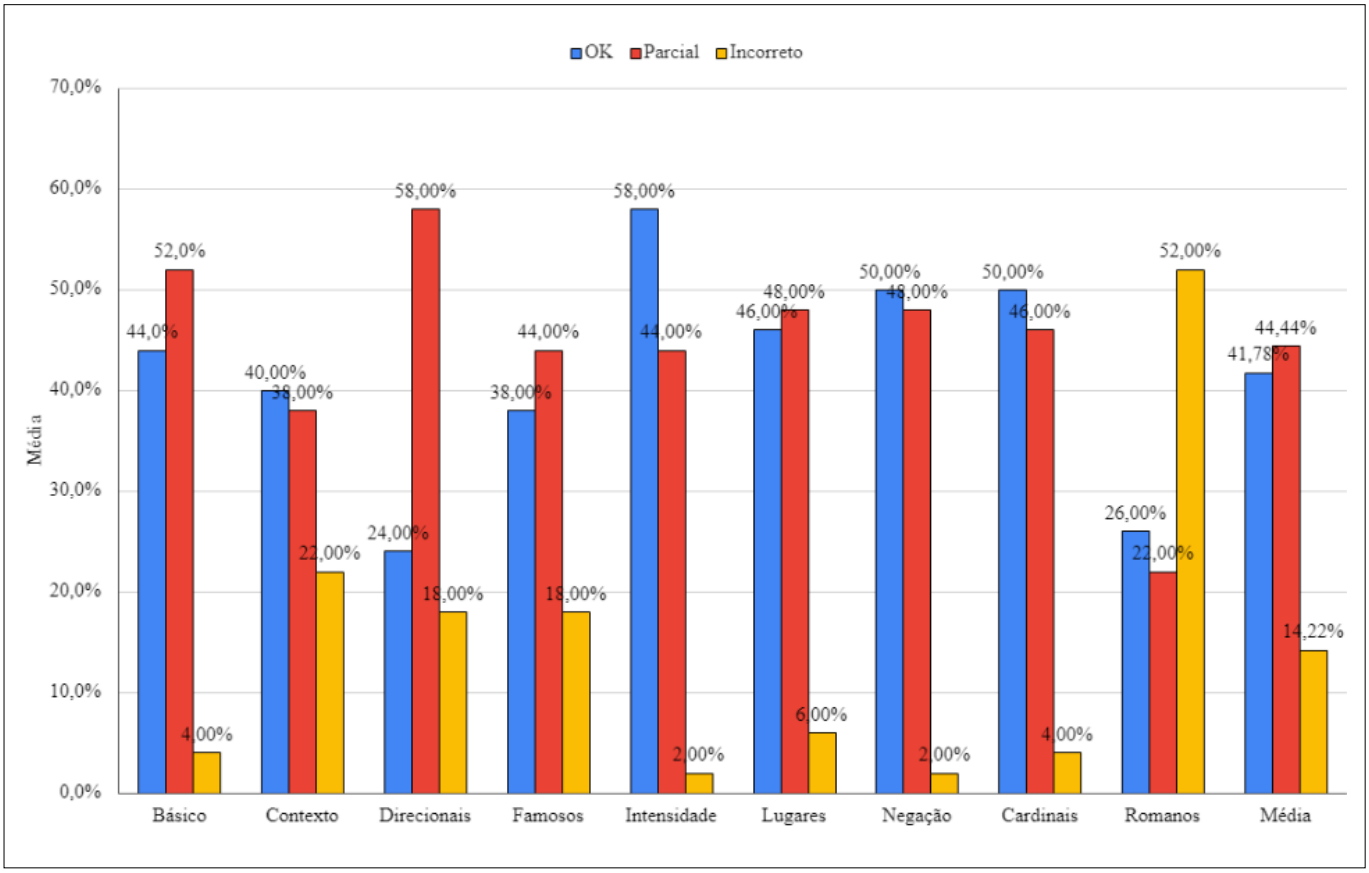
Na Figura 4, são exibidos os resultados obtidos pelo modelo BART. Observou-se uma melhoria nas traduções classificadas como OK, e uma redução, de aproximadamente 1 a 2 pontos percentuais, nas classificações Parcial e Incorreto, respectivamente. O subconjunto de sentenças envolvendo pessoas famosas foi o principal afetado. Esse baixo desempenho e alto custo de inferência tornam esse modelo pouco viável para o contexto do projeto.



ISSN 2447-9187

Figura 5 ▼
Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo *Transformer* básico.
Fonte: dados da pesquisa

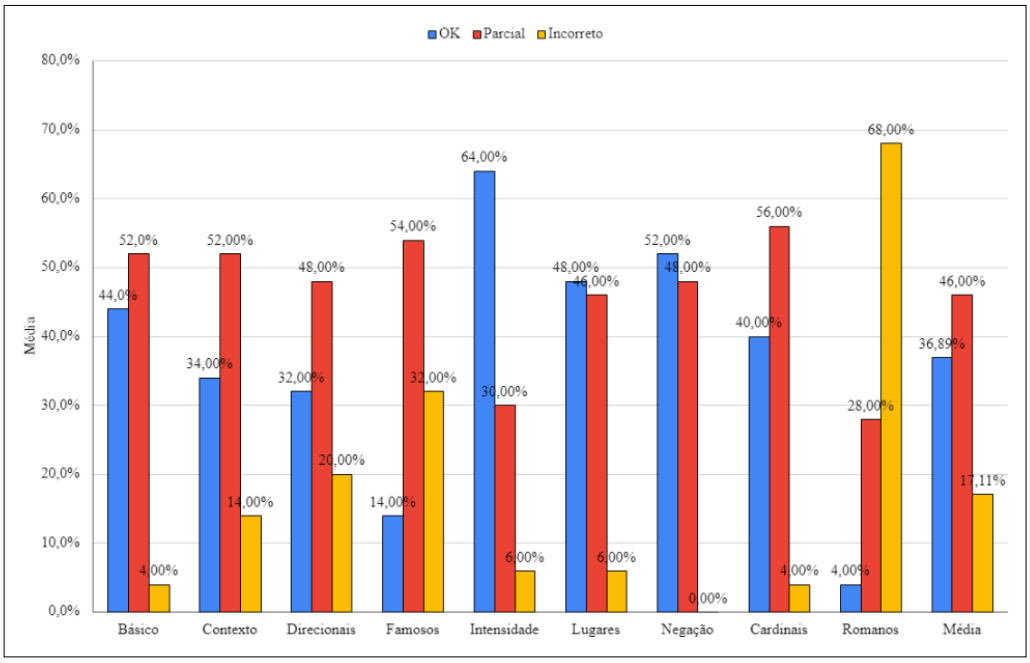
Os resultados obtidos com o modelo *Transformer* básico, apresentados na Figura 5, revelaram, pela primeira vez, uma melhora consistente de mais de 4,5 pontos percentuais nas traduções corretas. Esse progresso foi verificado em quase todos os subconjuntos de avaliação e sempre com uma migração das traduções parciais para traduções classificadas como OK.



Além da melhoria na qualidade da tradução, essa arquitetura apresenta um custo computacional próximo ao do modelo atualmente em produção e uma complexidade de integração relativamente baixa com o ecossistema do VLibras.

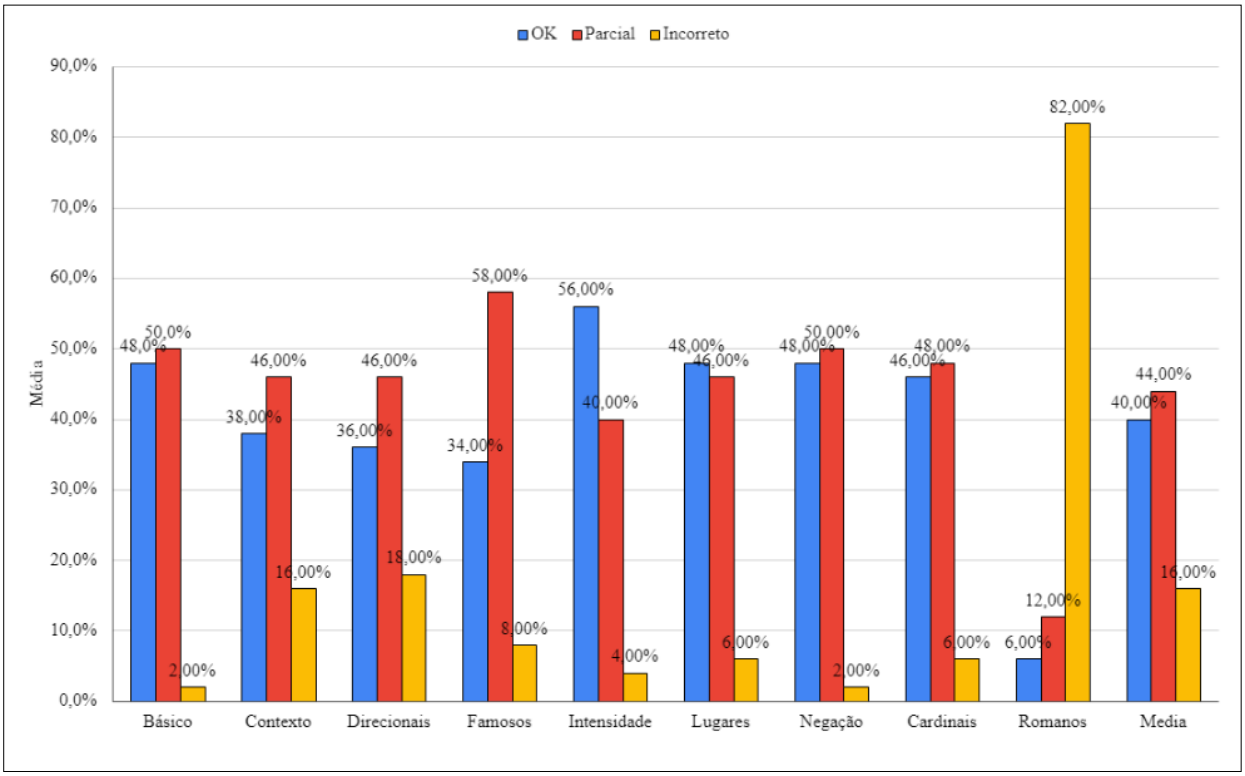
A Figura 6 apresenta os resultados obtidos com o modelo T5. Novamente, houve uma melhoria discreta nas traduções corretas, com aumento de aproximadamente 1,5 ponto percentual, enquanto as traduções incorretas apresentaram piora superior a 2%. As categorias de sentenças mais afetadas foram aquelas com indicativo de intensidade (melhora de 10%) e sentenças envolvendo pessoas famosas (piora de 12%).

Figura 6 ▶
Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo T5.
Fonte: dados da pesquisa



Após a conclusão do primeiro ciclo de experimentos com os modelos originais, duas variações do modelo T5 também foram analisadas. A Figura 7 mostra os resultados obtidos com a primeira variação, o modelo ByT5. A principal diferença entre essas subvariações é o processo de tokenização utilizado: enquanto o T5 utiliza uma abordagem de subpalavras baseada na SentencePiece (Kudo; Richardson, 2018), o ByT5 adota uma tokenização baseada em bytes (ou caracteres *unicodes*). Essa variação apresentou uma melhora significativa na maioria dos subconjuntos (exceto no subconjunto de números romanos), com aumento de aproximadamente 5% nas traduções corretas e uma redução de 1,5% nas traduções incorretas.

Figura 7 ▼
Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo ByT5.
Fonte: dados da pesquisa

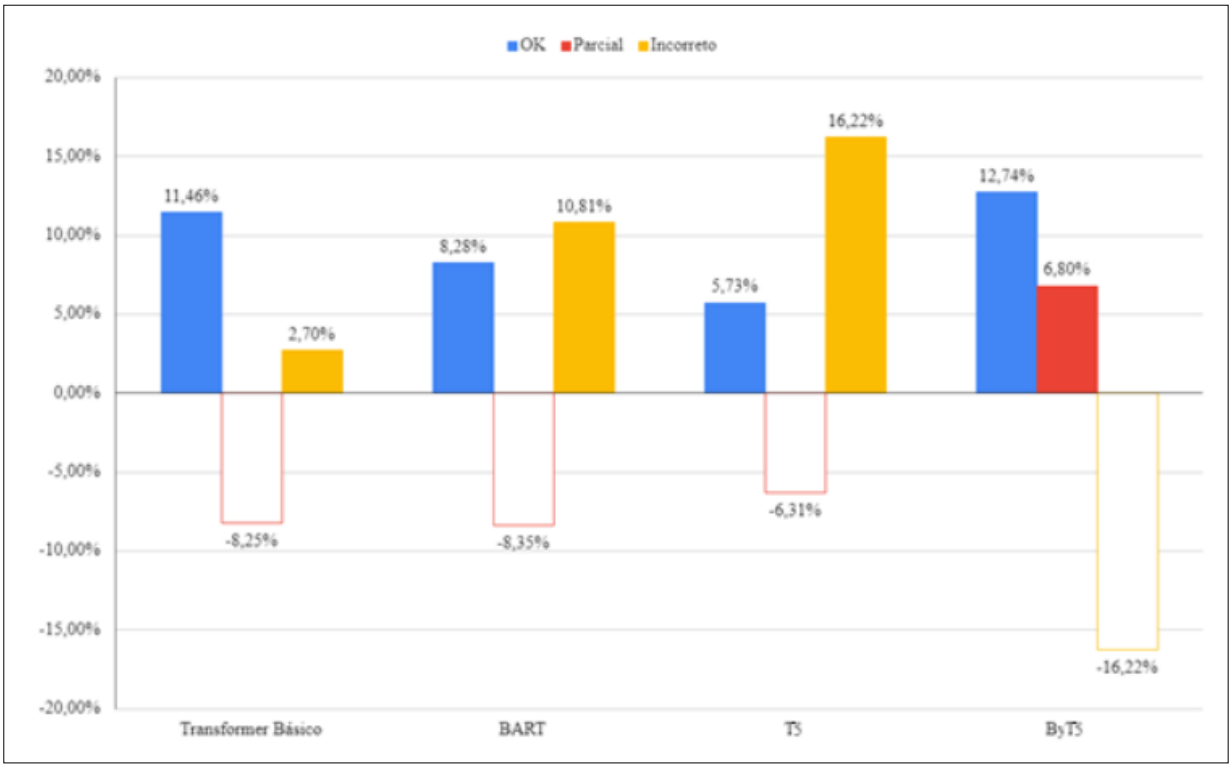


Outra variação testada foi a utilização do modelo *Transformer* básico em combinação com a técnica de *Back Translation*. Essa técnica, que realiza uma tradução adicional inversa durante a fase de treinamento, mostrou-se promissora, proporcionando uma melhoria média de quase 6% nas traduções corretas e redução discreta nas traduções incorretas. Observou-se, no entanto, uma inversão entre os percentuais obtidos de traduções corretas e traduções parcialmente corretas, além de piora discreta em algumas categorias de sentenças, exceto nas sentenças com algarismos romanos, cuja tradução piorou 4 pontos percentuais.

4.3 Discussão dos resultados

Na Figura 8 são resumidos os resultados dos experimentos. Observa-se que dois modelos candidatos apresentaram resultados superiores ao modelo de referência (LigthConv), com o modelo ByT5 alcançando as melhores médias, com um aumento percentual de 12,74% nas traduções perfeitas e diminuição das traduções incorretas em 16,22%. Em segundo lugar, destacou-se a arquitetura *Transformer* básico, com um aumento percentual de 11,46% para traduções corretas, embora as traduções incorretas tenham aumentado em 2,70%. Cada um desses modelos, entretanto, possui pontos positivos e negativos que precisam ser considerados.

Figura 8 ▼
Consolidação dos resultados obtidos pelos modelos avaliados em relação ao modelo de referência (LightConv).
Fonte: dados da pesquisa



A arquitetura BART foi descartada para o cenário em questão, pois, apesar de apresentar resultados superiores aos do modelo em produção, seu alto custo de inferência torna-a inviável para um ambiente de produção comercial com milhões de acessos mensais.

A arquitetura *Transformer* básico é considerada uma das mais eficientes em termos computacionais, por ser econômica e de fácil integração com sistemas como o VLibras. Além disso, permite um treinamento mais rápido, configurando-se como uma alternativa

atraente para projetos que requerem desempenho e eficiência computacional. Embora o custo computacional e a facilidade de integração sejam aspectos vantajosos, a precisão e a qualidade do modelo treinado com essa arquitetura tendem a ser menos competitivas, pois tal modelo não faz uso de modelos pré-treinados. Diante dessas considerações, entende-se que a arquitetura *Transformer* básico é uma opção viável para uso em produção.

A arquitetura T5 apresentou resultados mistos, com um aumento nas traduções perfeitas, mas também um aumento na taxa de traduções incorretas. O modelo ByT5, por sua vez, apresentou as melhores métricas computacionais e um desempenho superior em subconjuntos específicos de avaliação, como os subconjuntos de contexto e direcionalidade. Essa arquitetura é de alto custo computacional, aspecto que, no entanto, não representa um impeditivo para sua aplicação, especialmente após ser submetida a um processo automático de otimização e simplificação, usando-se a biblioteca FastT5[11]. Esse processo, embora viável, pode resultar em uma maior complexidade de integração. Mesmo com ressalva, a arquitetura ByT5 também foi considerada uma opção viável para o sistema VLibras.

[11] Disponível em: https://github.com/Ki6an/fastT5. Acesso em: 28 dez. 2023.

A arquitetura ByT5 também se mostrou vantajosa quando avaliada em um conjunto de dados que inclui conteúdos de páginas da internet com caráter institucional e/ou de serviços públicos (ver Tabela 1). Mesmo sem a inclusão de exemplos desses dados no processo de treinamento, a arquitetura ByT5 apresentou um ganho de 7,41 pontos percentuais na métrica BLEU, sugerindo uma melhor capacidade de generalização em comparação ao modelo atualmente em uso no VLibras.

Tabela 1 ▶
Comparativo da métrica BLEU entre o modelo de referência (LightConv) e a melhor arquitetura prospectada (ByT5).
Fonte: dados da pesquisa

Sentenças	LightConv	ByT5	Variação
Frases básicas	46,55	58,09	+11,54
Cardinais	72,51	71,69	-0,82
Contexto	54,40	50,50	-3,9
Direcionalidade	19,49	26,45	+6,96
Famosos	38,31	48,75	+10,44
Intensidade	45,13	48,78	+3,65
Lugares	47,46	56,09	+8,63
Negação	57,37	58,78	+1,41
Romanos	69,52	73,27	+3,75
Genéricas (sites)	25,38	32,79	+7,41
Média	47,61	52,52	+4,91

5 Conclusões

Esta pesquisa teve como objetivo apresentar um estudo comparativo de modelos neurais modernos, potencialmente aplicáveis na evolução do componente de tradução automática da Suíte VLibras. A plataforma em questão trata do processo de tradução do tipo texto para glosa entre a Língua Portuguesa e a Língua Brasileira de Sinais (Libras). Para isso, foi realizado um levantamento bibliográfico dos principais métodos de tradução surgidos nos últimos anos, especialmente entre 2017 e 2023. Foi também elaborada uma breve descrição sobre a evolução desses métodos ao longo do período em questão.

Com base na pesquisa bibliográfica, identificou-se um conjunto de estudos que abordam o problema da tradução automática, com foco no desafio de tradução entre línguas faladas e línguas de sinais. Observou-se que o uso de arquiteturas fundamentadas em mecanismo de atenção, especialmente modelos baseados em *Transformers*, tem ganhado destaque devido às melhorias significativas na qualidade da tradução.

Constatou-se que as soluções baseadas em arquiteturas *Transformers* representam o estado da arte para quase todos os problemas de NLP e até mesmo para desafios de visão computacional com os *Vision Transformers* (Dosovitskiy *et al*., 2021), tornando-se o novo padrão da indústria para diversos problemas práticos. Nesse contexto, os experimentos foram direcionados para avaliar a aplicabilidade dessas arquiteturas em contextos de processamento de linguagem natural com poucos recursos (*low-resources* NLP), que é o caso das línguas de sinais.

Este estudo mostrou, por meio de alguns experimentos, que a adoção de uma dessas arquiteturas viáveis (*Transformer* básico ou ByT5) pode contribuir para aumentar a precisão e a qualidade do componente de tradução da Suíte VLibras, proporcionando um aumento máximo de até 12,73% nas traduções perfeitas, redução de 16,21% nas traduções incorretas e uma melhoria média de 10,31% na métrica BLEU.

Os resultados indicam que os modelos baseados na arquitetura *Transformer* são promissores e podem ser considerados para uma eventual substituição do modelo neural usado na abordagem híbrida da Suíte VLibras, podendo simplificar o componente tradutor da plataforma para uma abordagem puramente neural.

Como continuidade deste estudo, pretende-se incluir outros conjuntos de dados na fase de avaliação da qualidade de tradução, abrangendo frases e construções presentes em sites governamentais e privados, poesias, conteúdos literários e outros contextos, visando aprimorar a avaliação do grau de generalização dos modelos avaliados. Outra possibilidade de estudo futuro é investigar se os modelos *Transformer* básico e ByT5 também podem oferecer ganhos similares na qualidade de tradutores neurais usados em outras línguas de sinais.

Agradecimentos

Os autores gostariam de agradecer ao Laboratório de Aplicações de Vídeo Digital (LAVID) da Universidade Federal da Paraíba (UFPB) pelo suporte logístico e operacional para a realização dos experimentos e agradecer também pelo apoio financeiro do Ministério da Gestão da Inovação em Serviços Públicos (MGI) e do Ministério dos Direitos Humanos e da Cidadania (MDHC) do Governo Federal do Brasil, os quais tornaram esta pesquisa possível.

Financiamento

Essa pesquisa foi realizada dentro do escopo da TED 02/2022 firmada entre o Ministério dos Direitos Humanos e da Cidadania (MDHC) e a UFPB.

Conflito de interesses

Os autores declaram não haver conflito de interesses.

Nota

Esta pesquisa é parte da dissertação de mestrado de Renan Paiva Oliveira Costa, intitulada "Uma Investigação sobre a Aplicabilidade de Redes Transformers no Contexto de Tradução Automática para Língua Brasileira de Sinais", orientada pelo Prof. Dr. Tiago Maritan Ugulino de Araújo e defendida em fevereiro de 2024 no Programa de Pós-Graduação em Informática (PPGI) do Centro de Informática (CI) da UFPB, disponível em https://sigaa.ufpb.br/sigaa/public/programa/defesas.jsf?lc=pt_BR&id=1879.

Contribuições ao artigo

COSTA, R. P. O.; ARAUJO, T. M. U: concepção ou desenho do estudo/pesquisa; análise e/ou interpretação dos dados; revisão final com participação crítica e intelectual no manuscrito. **SILVA, D. R. B.; MOREIRA, S. M.:** concepção ou desenho do estudo/pesquisa e apoio na configuração e realização dos experimentos. **SOUZA, D. F. L.; COSTA, R. E. O.:** análise e/ou interpretação dos dados e revisão final com participação crítica e intelectual no manuscrito. Todos os autores participaram da escrita, discussão, leitura e aprovação da versão final do artigo.

Referências

ABUJAR, S.; MASSUM, A. K. M.; BHATTACHARYA, B.; DUTTA, S.; HOSSAIN, S. A. English to Bengali neural machine translation using global attention mechanism. *In*: TAVARES, J. M. R. S.; CHAKRABARTI, S.; BHATTACHARYA, A.; GHATAK, S. (ed.). **Emerging Technologies in Data Mining and Information Security**. Singapore: Springer, 2021. p. 359-369. (Lectures Notes in Networks and Systems, v. 164). DOI: https://doi.org/10.1007/978-981-15-9774-9_35.

AMIN, M.; HEFNY, H.; MOHAMMED, A. Sign language gloss translation using deep learning models. **International Journal of Advanced Computer Science and Applications (IJACSA)**, v. 12, n. 11, p. 686-692, 2021. DOI: https://dx.doi.org/10.14569/IJACSA.2021.0121178.

ANGELOVA, G.; AVRAMIDIS, E.; MÖLLER, S. Using neural machine translation methods for sign language translation. *In*: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: STUDENT RESEARCH WORKSHOP, 60., 2022, Dublin. **Proceedings** […]. Dublin: ACL, 2022. p. 273-284. DOI: https://doi.org/10.18653/v1/2022.acl-srw.21.

ARAÚJO, T. M. U. **Uma solução para geração automática de trilhas em língua brasileira de sinais em conteúdos multimídia**. 2012. Tese. (Doutorado em Automação e Sistemas) – Universidade Federal do Rio Grande do Norte, Natal, 2012. Disponível em: https://repositorio.ufrn.br/handle/123456789/15190. Acesso em: 22 dez. 2023.

ARVANITIS, N.; CONSTANTINOPOULOS, C.; KOSMOPOULOS, D. Translation of sign language glosses to text using sequence-to-sequence attention models. *In*: INTERNATIONAL CONFERENCE ON SIGNAL-IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS (SITIS), 15., 2019, Sorrento. **Proceedings** […]. Sorrento: IEEE, 2019. p. 296-302. DOI: https://doi.org/10.1109/SITIS.2019.00056.

CAMGOZ, N. C.; HADFIELD, S.; KOLLER, O.; NEY, H.; BOWDEN, R. Neural sign language translation. *In*: 2018 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2018, Salt Lake City. **Proceedings** […]. Salt Lake City: IEEE, 2018. p. 7784-7793. DOI: https://doi.org/10.1109/CVPR.2018.00812.

CAMGOZ, N. C.; KOLLER, O.; HADFIELD, S.; BOWDEN, R. Sign language transformers: joint end-to-end sign language recognition and translation. *In*: 2020 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2020, Seattle. **Proceedings** […]. Seattle: IEEE, 2020. p. 10023-10033. DOI: https://doi.org/10.1109/CVPR42600.2020.01004.

CORRÊA, Y.; CRUZ, C. R. (org.). **Língua brasileira de sinais e tecnologias digitais**. Porto Alegre: Penso, 2019.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHEIR, T.; DEHGANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. An image is worth 16x16 words: Transformers for image recognition at scale. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2021), 9., 2021, *on-line*. **Proceedings** […]. Viena: OpenReview, 2021. Disponível em: https://openreview.net/forum?id=YicbFdNTTy. Acesso em: 2 jan. 2024.

DRYER, M. S.; HASPELMATH, M. **The world atlas of language structures (WALS)**. 2013. Disponível em: https://wals.info/. Acesso em: 22 dez. 2023.

FAROOQ, U.; RAHIM, M. S. M.; SABIR, N.; HUSSAIN, A.; ABID, A. Advances in machine translation for sign language: approaches, limitations, and challenges. **Neural Computing and Applications**, v. 33, n. 21, p. 14357-14399, 2021. DOI: https://doi.org/10.1007/s00521-021-06079-3.

GAGE, P. A new algorithm for data compression. **C Users Journal**, v. 12, n. 2, p. 23-38, 1994. Disponível em: https://dl.acm.org/doi/10.5555/177910.177914. Acesso em: 28 dez. 2023.

GÓMEZ, S. E.; MCGILL, E.; SAGGION, H. Syntax-aware transformers for neural machine translation: The case of text to sign gloss translation. *In*: WORKSHOP ON BUILDING AND USING COMPARABLE CORPORA (BUCC 2021), 14., 2021, *on-line*. **Proceedings** [...]. [*S.l.*]: ACL Anthology, 2021. p. 18-27. Disponível em: https://aclanthology.org/2021.bucc-1.4. Acesso em: 22 dez. 2023.

HAMED, H.; HELMY, A. M.; MOHAMMED, A. Holy quran-italian seq2seq machine translation with attention mechanism. *In*: INTERNATIONAL MOBILE, INTELLIGENT, AND UBIQUITOUS COMPUTING CONFERENCE (MIUCC), 2., 2022, Cairo. **Proceedings** […]. Cairo: IEEE, 2022. p. 11-20. DOI: https://doi.org/10.1109/MIUCC55081.2022.9781781.

HAQUE, R.; LIU, C.-H.; WAY, A. Recent advances of low-resource neural machine translation. **Machine Translation**, v. 35, p. 451-474, 2021. DOI: https://doi.org/10.1007/s10590-021-09281-1.

KHAN, M.; ULLAH, K.; ALHARBI, Y.; ALFERAIDI, A.; ALHARBI, T. S.; YADAV, K.; ALSHARABI, N.; AHMAD, A. Understanding the research challenges in low-resource

language and linking bilingual news articles in multilingual news archive. **Applied Sciences**, v. 13, n. 15, 8566, 2023. DOI: https://doi.org/10.3390/app13158566.

KOEHN, P.; KNOWLES, R. Six challenges for neural machine translation. *In*: WORKSHOP ON NEURAL MACHINE TRANSLATION, 1., 2017. Vancouver. **Proceedings** […]. Vancouver: ACL, 2017. p. 28-39. DOI: https://doi.org/10.18653/v1/W17-3204.

KUDO, T.; RICHARDSON, J. SentencePiece: a simple and language independent subword tokenizer and detokenizer for neural text processing. *In*: 2018 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2018, Brussels. **Proceedings** […]. Brussels: ACL, 2018. p. 66-71. DOI: https://doi.org/10.18653/v1/D18-2012.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. **Soviet Physics-Doklady**, v. 10, n. 8, p. 707-710, 1966. Disponível em: https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf. Acesso em: 22 dez. 2023.

LEWIS, M. P. **Ethnologue**: languages of the world. 17. ed. Dallas: Sil International, 2014.

LIU, Z.; WINATA, G. I.; FUNG, P. Continual mixed-language pre-training for extremely low-resource neural machine translation. *In*: FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL-IJCNLP 2021), 2021, *on-line*. **Proceedings** […]. [*S.l.*]: ACL, 2021. p. 2706–2718. DOI: https://doi.org/10.18653/v1/2021.findings-acl.239.

MOHAMED, A.; HEFNY, H.; AMIN, M. A deep learning approach for gloss sign language translation using Transformer. **Journal of Computing and Communication**, v. 1, n. 2, p. 1-8, 2022. DOI: https://dx.doi.org/10.21608/jocc.2022.254979.

OLIVEIRA, C. C. M.; RÊGO, T. G.; LIMA, M. A. C. B.; ARAÚJO, T. M. U. Analysis of rule-based machine translation and neural machine translation approaches for translating Portuguese to LIBRAS. *In*: BRAZILLIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 25., 2019, Rio de Janeiro. **Proceedings** […]. Rio de Janeiro: ACM, 2019. p. 117-124. DOI: https://doi.org/10.1145/3323503.3360305.

ORTEGA, J. E.; MAMANI, R. C.; CHO, K. Neural machine translation with a polysynthetic low resource language. **Machine Translation**, v. 34, n. 4, p. 325-346, 2020. DOI: https://dx.doi.org/10.1007/s10590-020-09255-9.

PAPINENI, K.; ROUKOS, S.; WARD, T.; ZHU, W.-J. BLEU: a method for automatic evaluation of machine translation. *In*: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 40., 2002, Philadelphia. **Proceedings** […]. Philadelphia: ACM, 2002. p. 311-318. DOI: https://doi.org/10.3115/1073083.1073135.

QUADROS, R. M. Efeitos de modalidade de língua: as línguas de sinais. **ETD – Educação Temática Digital**, v. 7, n. 2, p. 168-178, 2006. DOI: https://doi.org/10.20396/etd.v7i2.801.

RANATHUNGA, S.; LEE, E.-S. A.; SKENDULI, M. P.; SHEKHAR, R.; ALAM, M.; KAUR, R. Neural machine translation for low-resource languages: a survey. **ACM Computing Surveys**, v. 55, n. 11, 229, 2023. DOI: https://doi.org/10.1145/3567592.

RIVERA-TRIGUEROS, I.; OLVERA-LOBO, M.-D.; GUTIÉRREZ-ARTACHO, J. Overview of machine translation development. *In*: KHOSROW-POUR, M. (ed.). **Encyclopedia of Information Science and Technology**. 5. ed. Hershey: IGI Global, 2021. p. 874-886. DOI: https://dx.doi.org/10.4018/978-1-7998-3479-3.ch060.

SAUNDERS, B.; CAMGOZ, N. C.; BOWDEN, R. Progressive transformers for end-to-end sign language production. *In*: VEDALDI, A.; BISCHOF, H.; BROX, T.; FRAHM, J.-M. (ed.). **Computer Vision – ECCV 2020**. Cham: Springer, 2020. p. 687-705. (Lecture Notes in Computer Science, v. 12356). DOI: https://doi.org/10.1007/978-3-030-58621-8_40.

SHAZEER, N.; MIRHOSEINI, A.; MAZIARZ, K.; DAVIS, A.; LE, Q.; HINTON, G.; DEAN, J. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2017), 2017, Toulon. **Proceedings** […]. Toulon: OpenReview, 2017. Disponível em: https://openreview.net/pdf?id=B1ckMDqlg. Acesso em: 2 jan. 2024.

SOUZA, M. F. N. S.; ARAÚJO, A. M. B.; SANDES, L. F. F.; FREITAS, D. A.; SOARES, W. D.; VIANNA, R. S. M.; SOUSA, A. A. D. Principais dificuldades e obstáculos enfrentados pela comunidade surda no acesso à saúde: uma revisão integrativa de literatura. **Revista CEFAC**, v. 19, n. 3, p. 395-405, 2017. DOI: https://doi.org/10.1590/1982-0216201719317116.

WAGNER FILHO, J. A.; WILKENS, R.; IDIART, M.; VILLAVICENCIO A. The brWaC corpus: a new open resource for Brazilian Portuguese. *In*: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC 2018), 11., 2018, Miyazaki. **Proceedings** […]. Miyazaki: ELRA, 2018. Disponível em: https://aclanthology.org/L18-1686. Acesso em: 22 dez. 2023.

WILCOX, S.; WILCOX, P. P. **Aprender a ver**. Rio de Janeiro: Arara Azul, 1997.

WU, F.; FAN, A.; BAEVSKI, A.; DAUPHIN, Y.; AULI, M. Pay less attention with lightweight and dynamic convolutions. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2019), 2019, New Orleans. **Proceedings** […]. New Orleans: OpenReview, 2019. Disponível em: https://openreview.net/forum?id=SkVhlh09tX. Acesso em: 2 jan. 2024.

YIN, K.; READ, J. Attention is all you sign: sign language translation with transformers. *In*: SIGN LANGUAGE RECOGNITION, TRANSLATION AND PRODUCTION (SLRTP), 2020, *on-line*. **Proceedings** […]. [*S.l.*]: [*s.n.*], 2020. Disponível em: https://www.slrtp.com/papers/extended_abstracts/SLRTP.EA.12.009.paper.pdf. Acesso em: 22 dez. 2023.

YONGLAN, L.; WENJIA, H. English-Chinese machine translation model based on bidirectional neural network with attention mechanism. **Journal of Sensors**, v. 2022, 5199248, 2022. DOI: https://doi.org/10.1155/2022/5199248.

ZHANG, X.; DUH, K. Approaching sign language gloss translation as a low-resource machine translation task. *In*: BIENNIAL MACHINE TRANSLATION SUMMIT, 18.; INTERNATIONAL WORKSHOP ON AUTOMATIC TRANSLATION FOR SIGNED AND SPOKEN LANGUAGES (AT4SSL), 1., 2021, *on-line*. **Proceedings** […]. [*S.l.*]: Association for Machine Translation in the Americas, 2021. p. 60-70. Disponível em: https://aclanthology.org/2021.mtsummit-at4ssl.7/. Acesso em: 22 dez. 2023.