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), Brasil
* autor correspondente

Resumo
Os modelos recentes de *Neural Machine Translation* (NMT) podem ser aplicados a idiomas e domínios de poucos recursos sem limitações. Alguns trabalhos investigam se novas técnicas de NMT também podem ser generalizadas para diferentes recursos no que diz respeito à disponibilidade de dados e recursos computacionais. Nesse contexto, o objetivo geral deste estudo foi explorar métodos modernos de NMT e analisar a sua potencial aplicabilidade em contextos de poucos recursos, como é o caso 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 obtidos foram comparados com aqueles atualmente fornecidos pela atual arquitetura LightConv, sendo utilizado o mesmo *corpus* bilíngue Português-LIBRAS de treinamento e validação com mais de 70.000 sentenças geradas por linguístas, um dos maiores desse tipo disponíveis no mundo. Os resultados indicam que a adoção de uma das duas arquiteturas de melhor desempenho (*Basic Transformer* ou *ByT5*) ajudaria a melhorar a precisão e a qualidade da tradução da Suite VLibras, com um aumento percentual de até 12,73% considerando a métrica BLEU.
Palavras-chave: acessibilidade; línguas de poucos recursos; tradução automática neural; língua de sinais; transformers.

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

Abstract
Recent Neural Machine Translation (NMT) models may apply to low-resource languages and domains without limitations. Some works investigate whether new NMT techniques can also be generalized to different resources regarding data availability and computational resources. In this context, the general objective of this study was to explore modern NMT methods and analyze their potential applicability in low-resource contexts, such as sign languages. For a better evaluation, we adapted and used some promising identified models in the machine translation component of the VLibras Suite, and the obtained results were compared with those currently provided by the current LightConv architecture, using the same Portuguese-LIBRAS bilingual training and validation corpus with over 70,000 sentences generated by linguists, one of the largest of its kind available in the world. The results indicate that adopting one of the two top-performing architectures (Basic Transformer or ByT5) would help increase the accuracy and quality of the VLibras Suite translation, with a percentage increase of up to 12.73% considering 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 através de língua escrita ou falada. Isso se deve, principalmente, ao fato de que a maioria dos surdos passam vários anos na escola, mas não consegue atingir proficiência na leitura e escrita da língua oral de seu país (Souza

et al., 2017).

O principal motivo para essa dificuldade é que os surdos comunicam-se, naturalmente, através de línguas de sinais (LS), sendo as línguas orais (LO) apenas uma espécie de segunda língua. Cada LS, por sua vez, é uma língua natural, com léxico e gramática próprias, desenvolvida por cada comunidade de surdos ao longo do tempo, assim como cada comunidade de ouvintes desenvolveu a 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 existam muitas similaridades entre todas essas línguas, cada país normalmente tem a sua própria, alguns até mais de uma (Quadros, 2006).

Para permitir o acesso adequado, o ideal, portanto, é que os conteúdos em línguas orais sejam traduzidos ou interpretados para a LS associada. Contudo, considerando o volume e o dinamismo de informações em alguns ambientes e plataformas, como por exemplo, na Web, fazer isso usando intérpretes humanos é uma tarefa inviável, mesmo se for considerado apenas o conteúdo que é adicionado diariamente na Internet. Para endereçar de forma pragmática essa questão, uma das abordagens 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 (Rivera-Trigueros; Olvera-Lobo; Gutiérrez-Artacho, 2021): 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 a Tradução Automática Neural (*Neural Machine Translation* - NMT).

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, que é geralmente baseada em Aprendizagem Profunda (ou DL, do inglês *Deep Learning*), normalmente utiliza bases de dados com exemplos de sentenças tanto na língua de origem quanto na língua de destino para aprender a realizar as traduções.

Existem mais de 7.000 idiomas falados em todo o mundo, mas desses, apenas cerca de 20 tem corpo (ou *corpus*[1]) de texto 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 os idiomas da Europa Ocidental e também o idioma japonês (Lewis *et al.*, 2014). Por outro lado, a maioria dos idiomas falados na Ásia e na África não possuem os dados de treinamento necessários para construir sistemas NLP precisos. Essas linguagens são chamadas de linguagens de baixos recursos (do inglês *low resources languages*[2]) (Khan *et al.*, 2023). Esse também é o caso da maioria das línguas de sinais, com uma quase que total inexistência de material oralizado natural (escrito ou falado) em LS e quase sempre com poucos *corpus* bilíngues e, quando existentes, apenas de pequeno porte, em geral produzidos por linguistas.

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 NLP para linguagens de poucos recursos, como as línguas de sinais (Haque; Liu; Way, 2021; Khan *et al.*, 2023:

- 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 tarefas específicas com muita precisão como, por exemplo, detecção de discurso de ódio. No entanto, a criação de conjuntos de dados anotados requer intervenção humana, rotulando exemplos de treinamento um por um, tornando o processo geralmente demorado e muito caro, dados os milhares de exemplos exigidos pelos

[1] Quando um corpus possui um conjunto de sentenças equivalentes em mais de uma língua são chamados *corpus* bilíngue. Os conteúdos padrão em várias línguas, como, por exemplo, a bíblia, são uma ótima referência para a construção de *corpus* bilíngues.

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

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 são posteriormente ajustados para tarefas específicas. Portanto, abordagens para contornar a falta de conjuntos de dados não rotulados também se tornam muito importantes;
- Suporte a vários dialetos de um idioma: os idiomas que possuem vários dialetos também são 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 árabe estão em árabe padrão moderno. No entanto, para uma sensação humana ao interagir com assistentes de voz ou bate-papo para uso diário, ele é considerado muito formal para muitos falantes de árabe. Assim, os dialetos de suporte tornam-se necessários para casos de uso prático.

Algumas pesquisas recentes de NLP com poucos recursos (*low-resources* NLP) buscam personalizar soluções de NLP existentes baseadas em línguas mais ricas para idiomas e domínios com carência de dados. A premissa é que os modelos modernos de NLP podem ser igualmente aplicáveis tanto para linguagens de poucos recursos quanto para línguas com abundância de dados e, possivelmente, podem ser generalizadas com sucesso para cenários com diferentes níveis de recursos, tanto em termos de disponibilidade de dados e quanto em disponibilidade de recursos computacionais.

Neste contexto, o foco deste trabalho é avaliar alguns métodos em evidência baseados em *deep learning* e usados em tradução automática de *low resources languages* e analisar a sua potencial aplicabilidade para a tradução de línguas de sinais. O objetivo principal é identificar quais modelos, dentre os considerados, podem se adequar melhor para a tradução de Português Brasileiro para glosas 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[3] e os resultados obtidos comparados com os fornecidos atualmente pela ferramenta para avaliar se as novas abordagens podem representar alternativas para a melhoria na qualidade da tradução Português-LIBRAS disponível hoje.

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 apresentados os modelos de DL selecionados para o estudo e os critérios de elegibilidade utilizados. Na Seção 4 é apresentado o processo de desenvolvimento e avaliação dos modelos candidatos e os principais resultados obtidos. Por fim, as conclusões são apresentadas na Seção 5.

2 Trabalhos relacionados

Nesta seção são apresentados alguns trabalhos relacionados com o estudo alvo desta pesquisa e os modelos utilizados e/ou analisados nos mesmos que foram pré-selecionados como candidatos para a avaliação proposta.

No contexto de modelos baseados puramente em redes neurais profundas, Shazeer *et al.* (2017) propuseram um mecanismo denominado *Mixture of Experts* (MoE) que consiste de um número do que os autores denominaram de *experts*, que se traduzem em um conjunto de redes neurais *feed-forward* combinados em uma rede de bloqueio que seleciona combinações esparsas dos ditos *experts* para processar cada entrada. O mecanismo MoE é aplicado intermediariamente a uma pilha de redes LSTM

[3] A Suíte VLibras (Araújo, 2012) é o resultado de uma parceria entre o Ministério de Planejamento, Desenvolvimento e Gestão (MP), através da Secretaria de Tecnologia da Informação (STI), e a Universidade Federal da Paraíba (UFPB), através do Laboratório de Aplicações de Vídeo Digital (LAVID). Ela 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, dentre eles os principais sites do Governo Brasileiro (brasil.gov.br), da Câmara dos Deputados (camara.leg.br) e do Senado Federal (senado.leg.br) e está presente na vida cotidiana da comunidade surda através de milhões de traduções mensais. Mais informações podem ser obtidas em http://www.vlibras.gov.br.

(*Long Short-Term Memory).* Entre outros avanços destacados por Shazeer *et al.* (2017), no contexto de aplicação a tarefa de tradução, os autores conseguiram valores de BLEU de 40,56% para tradução inglês-francês (En-Fr) usando a base WMT'14/En-Fr e 26,03% para tradução Inglês-Dinarmarquês (En-De) usando a base WMT'14/En-De.

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

Na linha de modelos que se utilizam dos mecanismos de atenção, verificou-se uma quantidade crescente de trabalhos que seguem tal metodologia. 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 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 de ASL (*American Sign Language*) para o inglês. Os autores utilizaram três diferentes funções de atenção para construção da solução. No mesmo tema, o trabalho de Amin, Hefny e Mohammed (2021) propõe uma abordagem bidirecional a partir de GRU (*Gated Recurrent Units*), LSTM e mecanismo de atenção. Os autores aplicaram o modelo para tradução bidirecional entre a língua inglesa e a língua 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).

Na linha de soluções completamente 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 derivados.

No contexto da tradução de línguas faladas para línguas de sinais, Camgoz *et al.* (2020) propuseram o uso de *transformers* para atacar o problema de tradução de texto para glosa. Yin e Read (2020) utilizam modelos baseados em *transformers* e *Spatial-Temporal Multi-Cue* (SMTC) para executar as tarefas de reconhecimento e tradução de sinais. Na mesma linha, 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 leva em consideração informações de dependência léxica para o processo de tradução. Outros trabalhos que focaram no problema de reconhecimento e tradução de sinais para texto e texto para glosa foram desenvolvidos por Angelova, Avramidis e Moller (2022) e Mohamed, Hefny e Amin (2022).

Alguns dos novos modelos *transformers* se apresentam como ótimos candidatos para aplicação no problema de tradução de texto para glosa e que ainda foram pouco explorados ou não foram testados. Tais arquiteturas vão ser tratadas de forma mais aprofundada na Seção 3 deste artigo, a exemplo dos modelos BERT (*Bidirectional Encoder Representations from Transformers*), BART (*Bidirectional and Auto-Regressive Transformer*) e T5 (*Text-to-Text Transfer Transformer*).

3 Seleção de modelos candidatos

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

3.1 Critérios de elegibilidade

Desde a introdução da arquitetura LightConv em 2019 (modelo atual do tradutor neural do VLibras), 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 tem seu estado-da-arte baseado nessas redes.

Nesse sentido, a revisão da literatura para a prospecção de modelos realizada durante esta pesquisa e descrita na Seção 2 teve como objetivo identificar quais os modelos foram mais aplicados e/ou referenciados em artigos recentes da área, publicados entre 2017 e 2023 relacionados com o tema em pauta, sobretudo "*low resource* NLP" e/ou "*sign language* NMT".

Para essa fase de experimentação, alguns critérios de inclusão e exclusão adicionais foram definidos para seleção dos modelos candidatos para uma avaliação mais detalhada. Assim, além do possível ganho potencial de qualidade na tradução, outros fatores também foram considerados ao escolher os 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

Dentre as muitas 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 retro citados, foram pré-selecionados e tiveram a viabilidade da experimentação verificada apenas os disponíveis no portal PapersWithCode[4], o qual reúne um acervo interessante de trabalhos de pesquisa reprodutíveis, disponibilizando tanto *datasets*, código fonte e *benchmarks* comparáveis, obtidos sobre *corpus* públicos relevantes.

Após essa fase de confirmação de viabilidade da experimentação, foram selecionados os seguintes modelos para serem avaliados de forma mais criteriosa:

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

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

Além disso, os novos modelos também são avaliados quanto ao seu custo computacional e de infraestrutura. Sendo assim, somente foram considerados modelos que consigam ser executados em ambientes (servidores) baseados apenas em CPUs. Esse talvez seja 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 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. Para o contexto dessa avaliação, um modelo candidato será considerado inviável se seu tempo de inferência em CPU for superior a 2 segundos.

4 Avaliação dos modelos candidatos

Nesta seção é descrita como a avaliação experimental dos modelos selecionados foi planejada e realizada assim como a apresentação e a discussão dos resultados obtidos.

4.1 Planejamento de experimentos

As subseções seguintes descrevem como a avaliação experimental proposta para o estudo foi planejada e conduzida, 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 de forma direta nos componentes de tradução do VLibras. Os resultados obtidos são comparados

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

com os produzidos pela versão atual do tradutor híbrido do VLibras, o qual usa o modelo LightConv do *framework* Fairseq[5].

Para permitir a comparação com os resultados já disponíveis do VLibras são usados nos experimentos, os mesmos *corpus* de treinamento e validação e também calculada as mesmas métricas de avaliação. O *corpus* de treinamento e validação utilizado 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. As frases e traduções foram desenvolvidas manualmente por linguistas e intérpretes e, atualmente, é um dos maiores *corpus* desse tipo disponível no mundo.

Para facilitar a execução dessa etapa de experimentação foi utilizada a biblioteca de aprendizado profundo para NLP Hugging Face[6], que oferece acesso a vários modelos pré-treinados de processamento de linguagem natural como modelos de linguagem e tokenizadores. Esses modelos pré-treinados podem ser utilizados para tarefas comuns de NLP, como classificação de texto, extração de entidade, tradução automática, entre outras. A biblioteca também fornece ferramentas para treinar e personalizar modelos para tarefas específicas e facilita a integração com outras bibliotecas e *frameworks*. Essa biblioteca foi selecionada devido à disponibilidade de todos os modelos prospectados, por ser de fácil integração e estar em constante desenvolvimento, garantindo, assim, uma boa manutenção para o projeto.

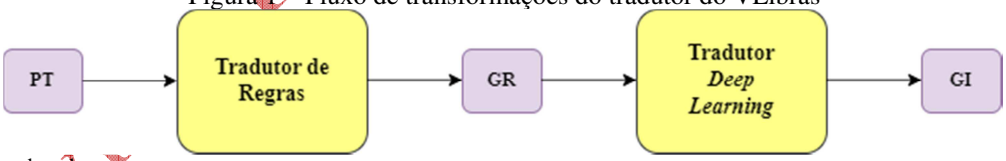
4.1.2 Arquitetura de tradução da Suíte VLibras

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) (modelo LightConv – Wu *et al.*, 2019). Nesse contexto, o tradutor de regras faz o papel de um componente de pré-processamento da sentença em português que, por sua, vez alimenta o modelo LightConv. Esse processo tem como objetivo normalizar a entrada e ajudar o modelo durante o treinamento. Essa etapa é fundamental em função do relativo baixo volume de dados disponíveis.

O fluxo de inferência do processo de tradução (Figura 1) é representado pelas etapas que uma frase em português passa até ser convertida em uma representação traduzida em glosa[7] 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.

Figura 1 – Fluxo de transformações do tradutor do VLibras



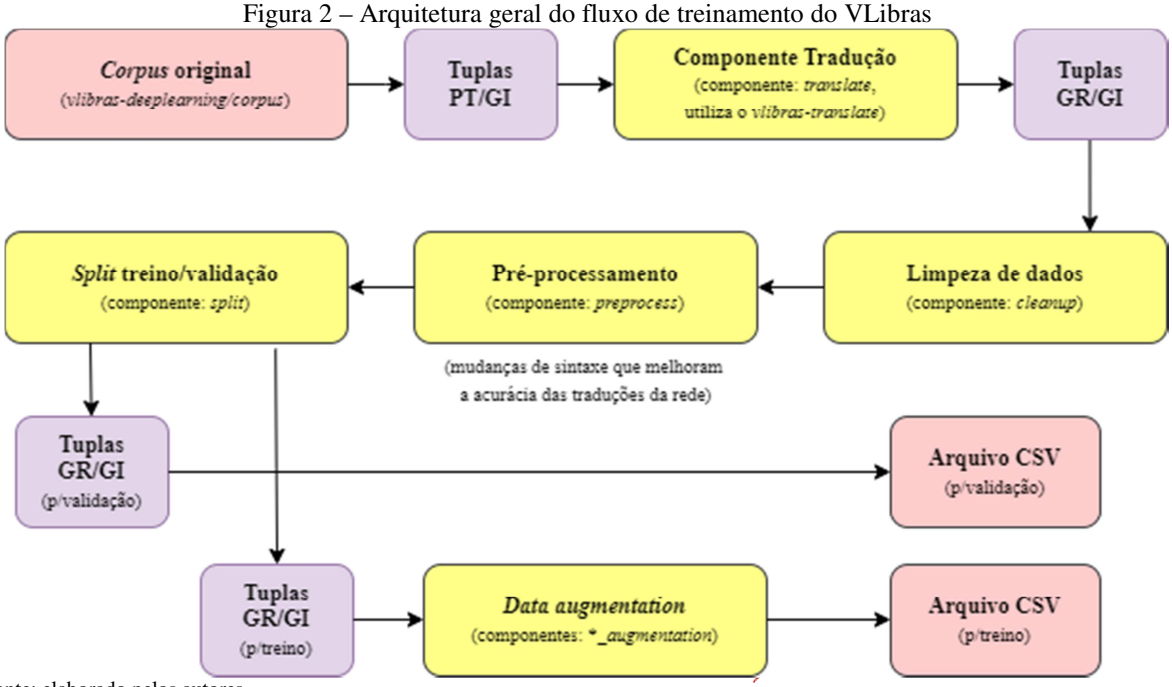
Fonte: elaborado pelos autores

O treinamento do tradutor *deep learning* do VLibras também possui um *pipeline* (Figura 2). O *corpus* bilíngue de treinamento, um conjunto de dados contendo diversas sentenças equivalente em português para glosa, é traduzido utilizando o tradutor de regras para glosa intermediária. Nesse processo, as sentenças passam por várias etapas de pré-processamentos. Em seguida, essas entradas são divididas em dois conjuntos distintos: um para ser usado no treinamento do modelo e outro para ser usado apenas na validação do processo de tradução. O conjunto de treinamento também passa por

[5] Fairseq é um kit de ferramentas de modelagem de sequência para treinar modelos personalizados para tradução, resumo e outras tarefas de geração de texto produzido pela Meta.
[6] Disponível em: https://huggingface.co. Acesso em 27 dez. 2023.
[7] Glosa são palavras de uma determinada língua oral grafadas com letras maiúsculas que representam sinais manuais de sentido próximo. Wilcox e Wilcox (1997) definem glosa como sendo uma tradução simplificada de morfemas da língua sinalizada para morfemas de uma língua oral.
[8] Frase simplificada e em estado intermediário para facilitar a tradução *deep learning*

um processo de *data augmentation*[9] para ampliar a ocorrência de palavras raras em seu conjunto de sentenças.

Figura 2 – Arquitetura geral do fluxo de treinamento do VLibras



Fonte: elaborado pelos autores

Após o pré-processamento do *corpus*, esses dados passam por 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 serem utilizadas para o treinamento do modelo usado no tradutor *deep learning* atual do VLibras (modelo LightConv do *framework* Fairseq).

4.1.3 Métricas de interesse

Os modelos de *deep learning* são treinados em grandes bases de dados denominadas *corpus*[11], sendo necessário realizar a avaliação automática dessas traduções, visando medir a eficiência do modelo de tradução automática. Até o presente momento, a métrica de avaliação mais usada para avaliação de tradução automática é conhecida como BLEU, a qual foi proposta pela primeira vez por Papineni *et al.* (2002). A estratégia da métrica BLEU é 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 onde múltiplas avaliações são necessárias.

O resultado da métrica BLEU é normalmente expressado como um número entre 0 e 1, onde 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. A métrica BLEU tem se mostrado eficiente em indicar o desempenho de modelos de tradução automática.

Além da métrica BLEU, também é possível avaliar tradução automática utilizando medidas de similaridade. A distância de Levenshtein (Levenshtein, 1966), também conhecida como distância de edição, é uma medida de similaridade entre duas *strings* ou sequências de caracteres. Ela é baseada 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. Ela calcula a distância de

[9] *Data augmentation* é um recurso muito aplicado em *low-resource NLP* para ampliar, sinteticamente e através de técnicas específicas, a quantidade de sentenças em *corpus* usados em treinamento de modelos neurais.
[10] *Byte pair encoding* (BPE) é um método de tokenização caracterizado por representar um texto com o menor número de bytes.
[11] Coleção de documentos ou textos escritos em determinada língua.

Levenshtein de forma 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, onde 0 indica que duas sentenças não possuem nenhuma palavra ou *token* em comum e 1 indica que as duas sentenças são idênticas.

A distância de Levenshtein normalizada é frequentemente usada como uma métrica de similaridade para *strings*. Ela é utilizada em diversas áreas como, por exemplo, detecção de plágios, processamento de linguagem natural, reconhecimento de fala, tradução automática, entre outros. A distância de Levenshtein normalizada é útil porque permite comparar *strings* de tamanhos diferentes de maneira justa. Além disso, é uma forma de contornar a desvantagem de distância de Levenshtein que é altamente sensível às diferenças de tamanho caso não seja normalizada.

Para avaliação do VLibras, uma tradução com distância de Levenshtein normalizada de valor igual a 1 é considerado correta, ou seja, é uma tradução perfeita. Uma distância de Levenshtein normalizada de valor menor do que 0,85 é considerada uma tradução incorreta e um valor maior do que 0,85 e menor do que 1 é considerado uma tradução parcialmente correta. Esses valores limiares foram estabelecidos empiricamente ao longo do desenvolvimento do componente de tradução do VLibras por meio de testes, avaliações e coleta da percepção de qualidade da tradução 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);
- Incorreto (similaridade menor que 0,85).

Portanto, a avaliação computacional usado atualmente no componente tradutor do VLibras usa duas métricas de interesse: uma métrica de tradução (BLEU) e uma métrica de similaridade (distância de Levenshtein normalizada). Para permitir uma comparação adequada, as mesmas métricas serão adotadas neste estudo.

4.1.4 Conjuntos de avaliação

Tão importante quanto à definição das métricas de interesse é a definição do conjunto de dados que será usado para avaliação do modelo, também chamado de conjunto de avaliação. Um conjunto de avaliação é um subconjunto do conjunto de dados de treinamento que é separado e usado apenas para avaliar o desempenho de um modelo de *deep learning*. Ele é usado para medir quão bem o modelo é capaz de generalizar para dados que ele nunca viu antes.

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

É importante notar que o conjunto de avaliação deve ser completamente separado do de treinamento e de validação, de forma que ele contenha dados que o modelo nunca viu antes. Isso é importante para evitar que o modelo "memorize" os dados de treinamento e validação, o que resultaria em uma superestimação do desempenho do modelo. Esse fenômeno é conhecido como sobreajuste dos dados de treinamento, do inglês *overfitting*.

Além disso, é importante avaliar o modelo em diferentes conjuntos de dados, tanto em termos de desempenho quanto de robustez. Para isso, além de avaliar o modelo em dados de teste, também é importante avaliar o desempenho do modelo com diferentes tipos de dados, como dados desbalanceados, dados incompletos, dados diferentes daqueles usados no treinamento e assim por diante. Dessa forma, é possível entender melhor como o modelo se comporta e identificar quaisquer problemas ou limitações.

No contexto do VLibras, essa avaliação é realizada sobre diferentes conjuntos de avaliação, que procuram modelar cenários e pontos críticos encontradas no processo de tradução de português para Libras, sendo projetados com a supervisão de especialistas em Libras:

- frases básicas;

- frases contendo referências de contexto;
- frases com referências direcionais;
- frases com sentido de negação;
- frases contendo nome 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 sobre cada um desses conjuntos antes de serem movidos para uma etapa de homologação e, finalmente, serem disponibilizados para o usuário final através dos componentes interativos da Suíte VLibras.

4.1.5 Configuração do ambiente

Para a realização dos experimentos foram utilizados dois ambientes de processamento para permitir uma paralelização de cada execução planejada, posto que cada ciclo de treinamento e validação durava cerca de 5 horas. Para a configuração de cada ambiente foi preciso instalar diversos módulos Python dos *frameworks* utilizados. Em seguida, foi baixado o código fonte do *pipeline* VLibras em produção e seus submódulos do serviço de versões GitLab, hospedado no Laboratório de Aplicações de Vídeo Digital da UFPB (LAVID). Os dois ambientes foram configurados de forma similar para o treinamento de cada modelo previsto, incluindo a adaptação do modelo e a integração do mesmo ao *pipeline* do VLibras.

Antes da execução dos experimentos foram conferidos os hiperparâmetros com os usados em produção e foram realizados testes de sanidade para aferir se ambos os ambientes forneciam resultados compatíveis e sincronizados. Adicionalmente, foram executados treinamentos exploratórios e comparados com os resultados do modelo atual e ajustes como versão de dependências, variáveis de ambiente, entre outros, foram realizados até que os resultados fossem equivalentes.

Durante essa fase, foi identificado que o modelo BERT não estava produzindo resultados adequados no contexto de tradução Português/Libras. De maneira geral, o modelo apresentou uma reprodutibilidade dos resultados dos artigos que o referenciavam bastante limitada, em especial para o problema de tradução automática, o que ocasionou o seu descarte para as fases seguintes de avaliação.

4.1.6 Realização dos experimentos

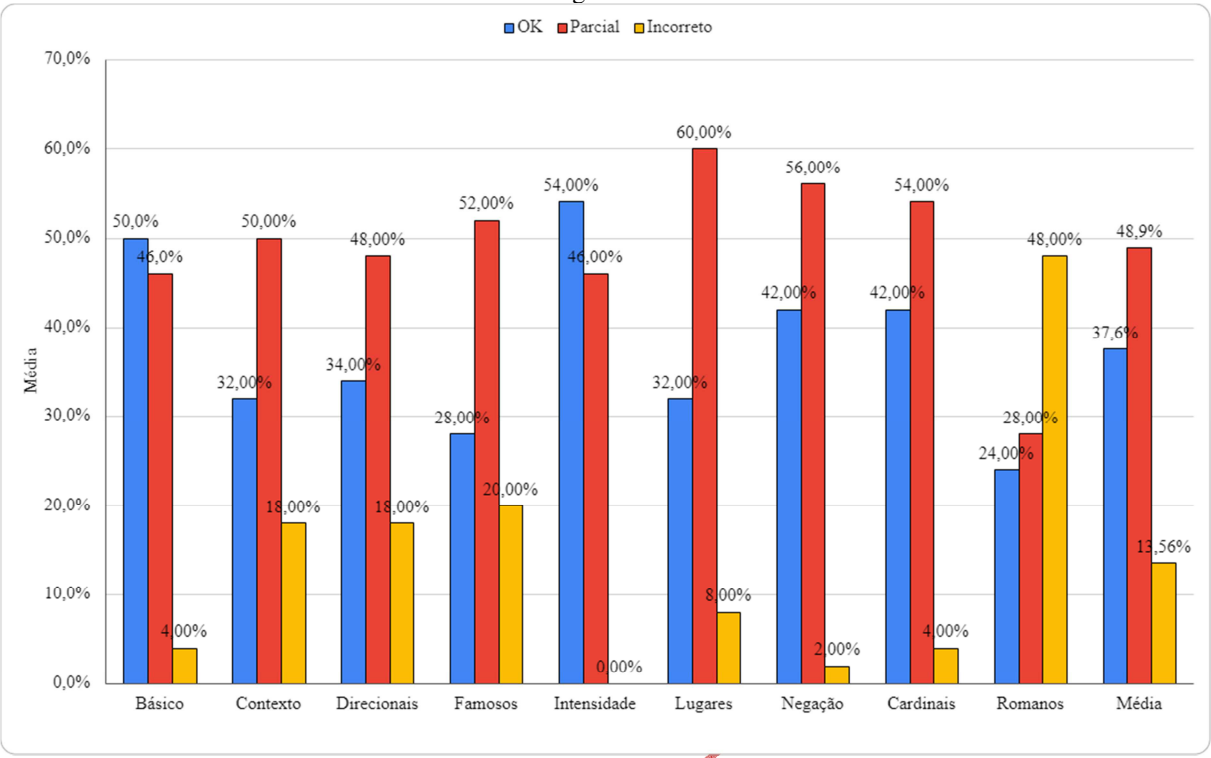
O treinamento e avaliação de cada experimento foram executados de forma paralela em um dos dois ambientes e os resultados calculados e consolidados para cada modelo e para cada subconjunto de sentenças de avaliação. O objetivo do primeiro ciclo de experimentos foi gerar uma pontuação de referência do modelo atual, LightConv. Para os ciclos seguintes foram utilizados os seguintes modelos: BART, *Transformer* básico, T5 e ByT5. Também foram combinadas com os modelos testados algumas técnicas 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

Nessa subseção são apresentados os resultados obtidos em cada ciclo de experimentos. Em geral, as tabelas de resultados possuem uma coluna para um dos nove subconjuntos de avaliação considerados e uma linha para cada uma das classificações possíveis para cada tradução (OK, Parcial e Incorreto). Os valores em cada célula trazem o percentual de resultados de cada classificação obtido em cada subconjunto pelo modelo/configuração sendo considerado.

O modelo que está em uso no VLibras atualmente, tradutor híbrido baseado no modelo LightConv, será usado como referência para avaliação de novos modelos, ou seja, será possível verificar se os novos modelos apresentam métricas computacionais melhores ou piores do que o modelo em produção (Figura 3).

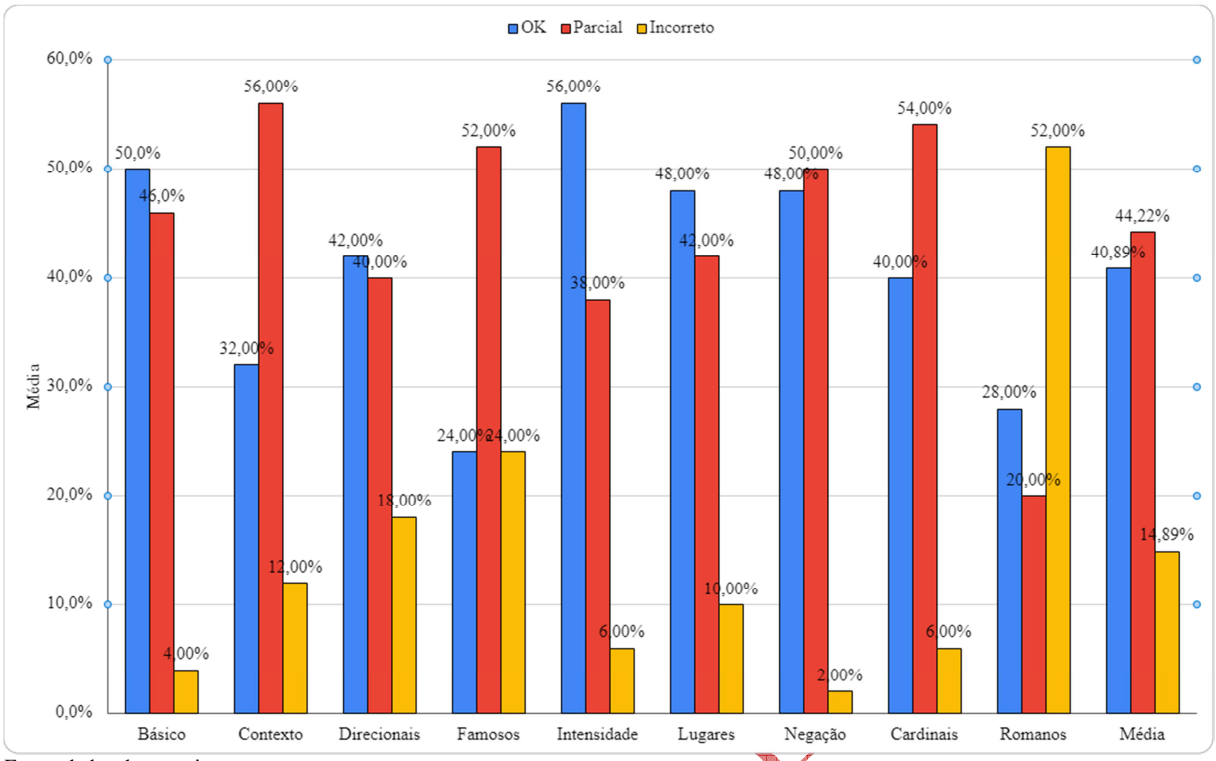
Figura 3 – Resultados da métrica de similaridade por tipo de sentença obtidos com modelo de referência LightConv



Fonte: dados da pesquisa

Na Figura 4 são exibidos os resultados obtidos pelo modelo BART. Houve uma melhora nas traduções OK e uma piora, na ordem entre 1 e 2 pontos percentuais, nos resultados parciais e incorretos, respectivamente. O principal subconjunto afetado foi o de sentenças com pessoas famosas. Esse baixo desempenho e seu alto custo de inferência torna esse modelo pouco viável para o contexto do projeto.

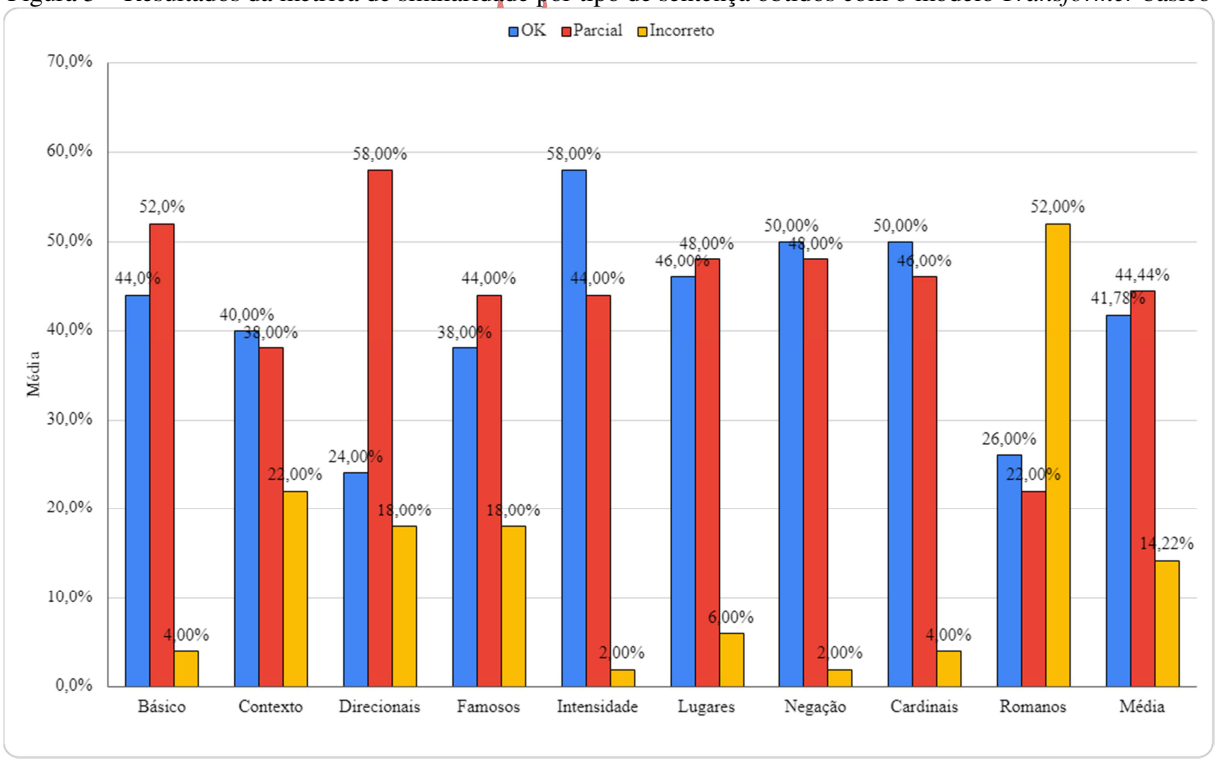
Figura 4 – Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo BART



Fonte: dados da pesquisa

Os resultados obtidos com o *Transformer* básico, por sua vez, podem ser visualizados na Figura 5. Pela primeira vez, um modelo conseguiu uma melhora consistente de mais de 4,5 pontos percentuais nas traduções corretas. Essa melhora foi observada em quase todos os subconjuntos de avaliação e sempre com uma migração de traduções parciais para traduções OK.

Figura 5 – Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo *Transformer* básico

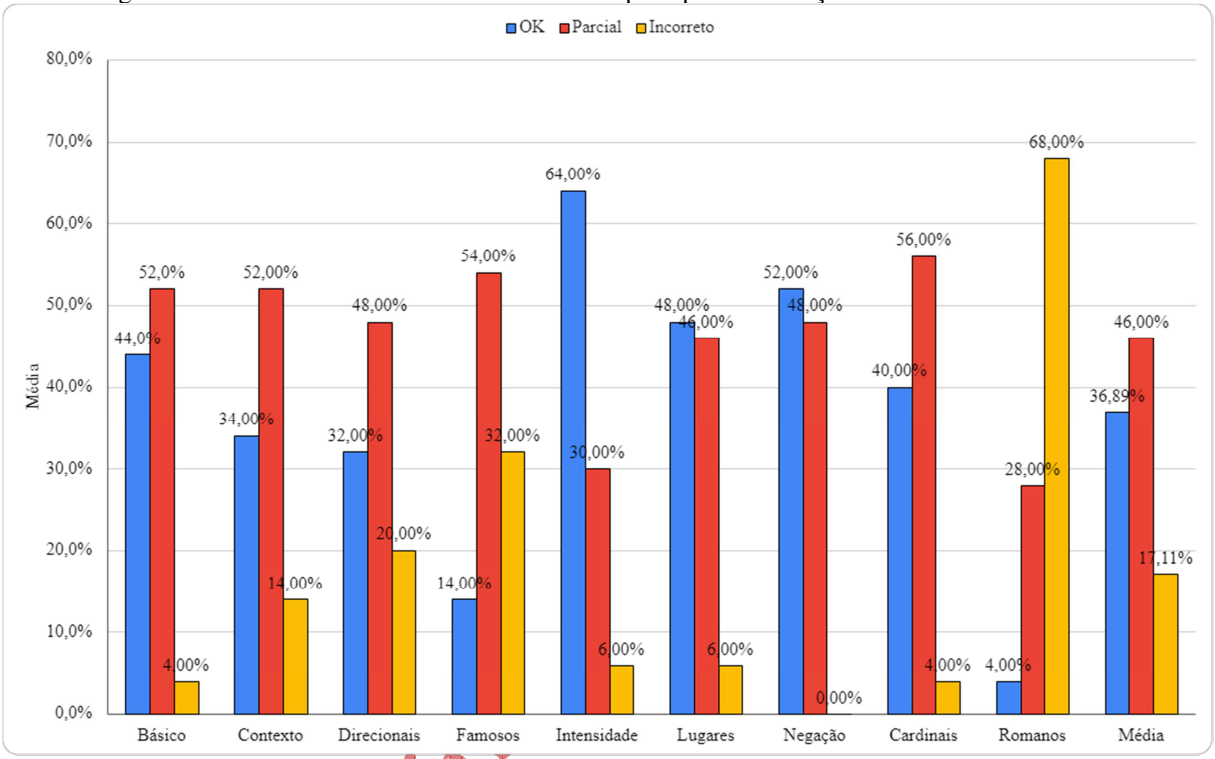


Fonte: dados da pesquisa

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

A Figura 6, por sua vez, apresenta os resultados obtidos com o modelo T5. Novamente houve uma melhora discreta das traduções corretas, de quase 1,5 pontos percentuais, enquanto que as traduções incorretas pioraram em mais de 2%. As categorias de sentenças mais afetadas foram as com indicativo de intensidade (melhora de 10%) e sentenças com pessoas famosas (piora de 12%).

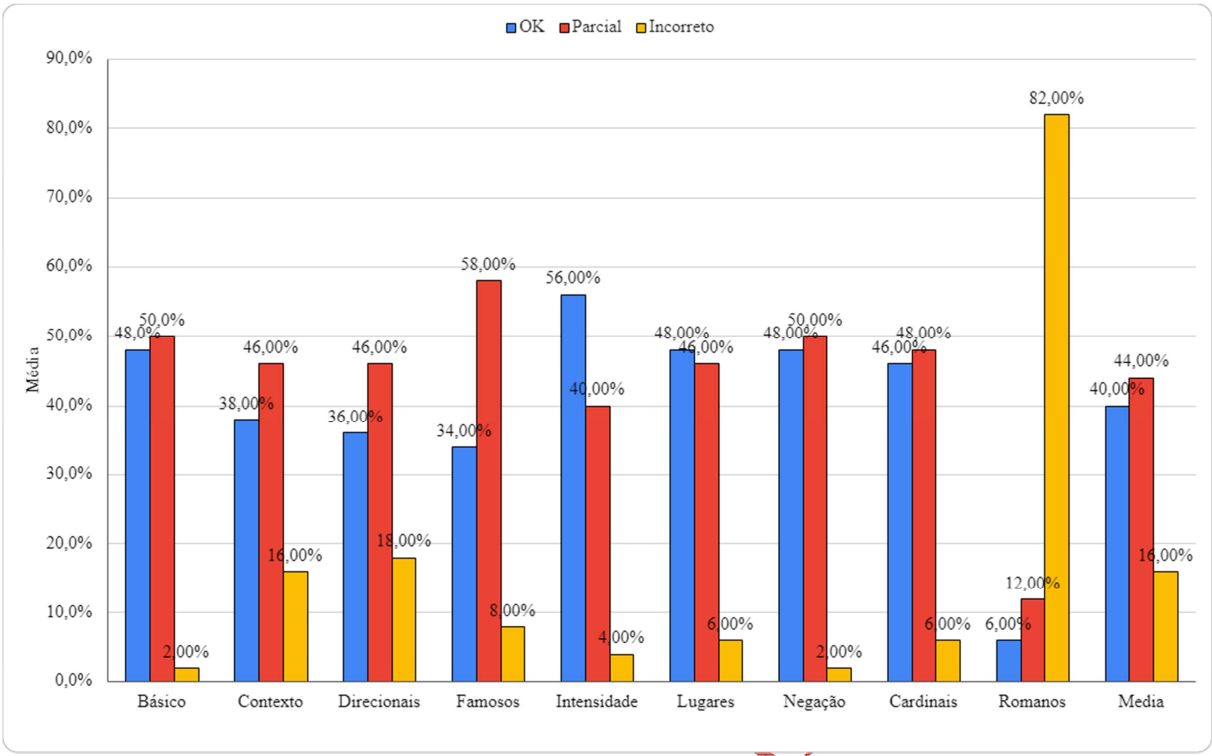
Figura 6 – Resultados da métrica de similaridade por tipo de sentença obtidos com o modelo T5



Fonte: dados da pesquisa

Após a conclusão do primeiro ciclo de experimentos com os modelos originais, duas variações do modelo T5 também foram consideradas. A Figura 7 indica os resultados obtidos com a primeira delas, o modelo ByT5. A principal diferença entre elas é o processo de tokenização utilizado, enquanto no T5 é usada uma abordagem de subpalavras baseada na SentencePiece (Kudo *et al.*, 2018), no ByT5 a tokenização é baseada em bytes (ou caracteres *unicodes*). Essa variação apresentou uma melhora significativa em quase todas as faixas de resultados (com exceção do subconjunto de número romanos), com um acréscimo de quase 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



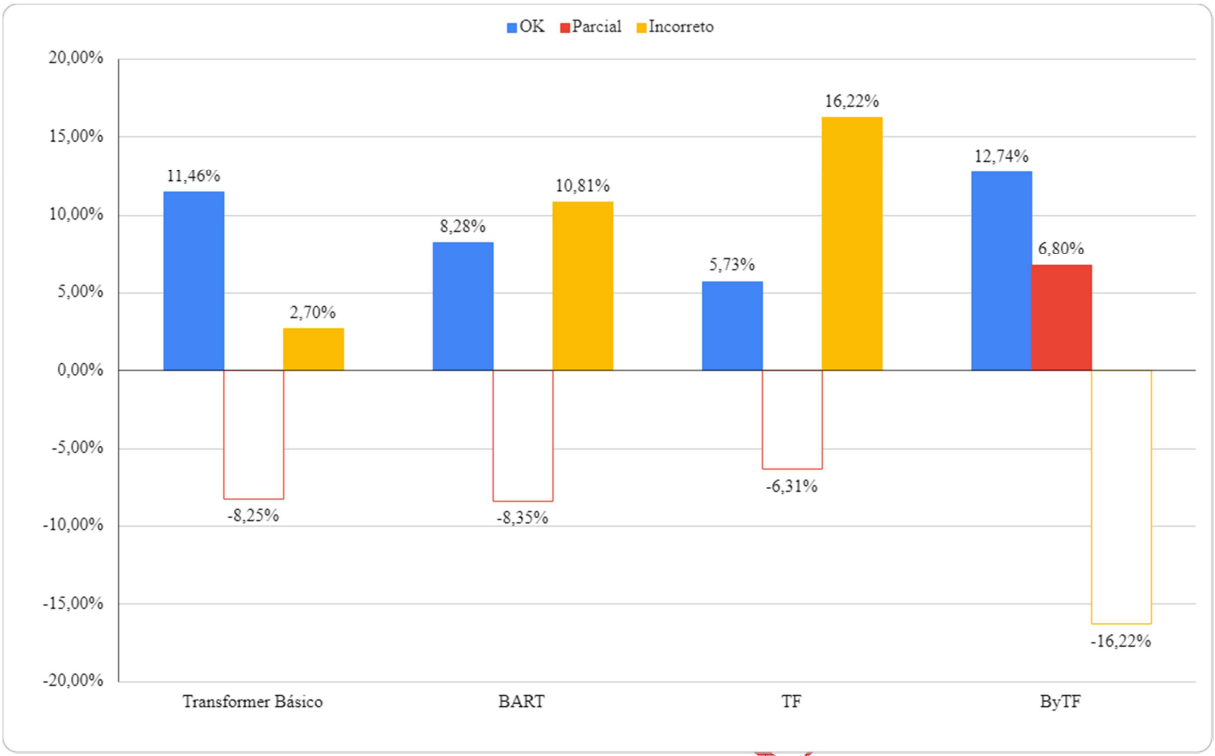
Fonte: dados da pesquisa

Outra variação testada foi a utilização do modelo *Transformer* básico de forma combinada com a técnica *Back Translation*. Essa técnica se baseia em fazer uma tradução adicional inversa durante a fase de treinamento e se mostrou promissora, com melhoria de quase 6% na média das traduções corretas e diminuição discreta na média de traduções incorretas. Um ponto de atenção, a ser investigado posteriormente, foi a inversão entre os percentuais obtidos de traduções corretas e traduções parcialmente corretas e algumas categorias de sentenças com piora discreta, com exceção da tradução de sentenças com algarismos romanos que piorou 4 pontos percentuais.

4.3 Discussão

Na Figura 8 é sumarizado os resultados dos experimentos. É possível perceber que dois modelos candidatos conseguiram melhores resultados do que o modelo de referência (LigthConv), sendo que o modelo ByT5 apresentou as melhores médias, obtendo um aumento percentual de 12,74% nas traduções perfeitas e diminuindo as traduções incorretas em 16,22%. Em segundo lugar, a arquitetura *Transformer Básico*, que obteve um aumento percentual de 11,46% para traduções corretas, porém as traduções incorretas aumentaram em 2,70%. No entanto, cada um desses modelos apresenta alguns prós e contras que precisam ser avaliados.

Figura 8 – Consolidação dos resultados obtidos pelos modelos avaliados em relação ao modelo de referência (LightConv)



Fonte: dados da pesquisa

A arquitetura BART, por sua vez, foi descartada para o cenário em pauta, pois, apesar de obter resultados superiores ao modelo em produção, a mesma apresenta um alto custo de inferência, sendo inviável em um cenário baseado em produção comercial com milhões de acessos mensais.

A arquitetura *Transformer* básico é considerada uma das arquiteturas mais eficientes do ponto de vista computacional, sendo uma das arquiteturas mais baratas e de fácil integração com sistemas como o VLibras. Além disso, ela também apresenta um treinamento mais rápido, tornando-a uma opção atraente para projetos que requerem desempenho e eficiência computacional. É importante destacar que, embora o custo computacional e a facilidade de integração sejam aspectos vantajosos, a precisão e a qualidade do modelo treinado com base na arquitetura *Transformer* básico tende a ser menos competitiva, pois não faz uso de modelos pré-treinados. Feitas tais considerações, entende-se que a arquitetura *Transformer* básico é considerada uma opção viável para uso em produção.

A arquitetura T5 apresentou métricas mistas, obtendo um aumento de traduções perfeitas, porém aumentando a taxa de traduções incorretas. Por sua vez, ByT5 apresentou as melhores métricas computacionais e um ganho expressivo em conjuntos de avaliação específicos, como o conjunto de contexto e direcionalidade. No entanto, ela possui um alto custo computacional, embora o mesmo não seja proibitivo, especialmente depois de passar por um processo automático de otimização e simplificação usando a biblioteca FastT5[12]. Esse processo é factível, porém pode resultar em uma maior complexidade de integração. Apesar desses aspectos, a ByT5 também foi considerada uma opção viável para um sistema como o VLibras.

A arquitetura ByT5 também se mostrou vantajosa quando avaliada sobre um conjunto de dados que compreende conteúdo de páginas da Internet de caráter institucional e/ou de serviços públicos (ver Tabela 1). Mesmo sem a introdução de exemplos de tais dados no processo de treinamento ByT5, houve um ganho de 7,41 pontos percentuais na métrica BLEU, indicando que esse modelo tem uma melhor capacidade de generalização do que o modelo atualmente em uso pelo VLibras.

Tabela 1 – Comparativo da métrica BLEU entre o modelo de referência (LightConv) e o melhor modelo prospectado (ByT5)

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

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

Fonte: dados da pesquisa

5. Conclusões

A presente pesquisa teve como objetivo apresentar um estudo comparativo de modelos neurais modernos e 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 tanto foi realizado um levantamento bibliográfico dos principais métodos de tradução que surgiram nos últimos anos, em especial no interstício entre 2017 e 2023. Uma breve descrição sobre a evolução dos métodos de tradução desenvolvidos ao longo deste período também foi realizada

A partir do levantamento bibliográfico, foi identificado um conjunto de trabalhos que endereçam o problema de tradução automática, em especial trabalhos com foco no problema de tradução entre línguas faladas e línguas de sinais. Foi observado que o uso de arquiteturas fundamentadas em mecanismo de atenção e, em especial, os modelos baseados em *Transformers* ganharam grande relevância nos últimos anos, visto as melhorias na qualidade de tradução apresentadas por tais métodos.

Foi constatado que as soluções baseadas em arquiteturas *Transformers* são o estado da arte para praticamente todos os problemas de NLP e, até mesmo, para problemas de visão computacional através dos *Vision Transformers* (Dosovitskiy *et al.*, 2020), sendo o novo padrão da indústria para vários problemas práticos. Nesse contexto, os experimentos foram focados em avaliar se tal adequação poderia também ser aplicada em contextos de *low-resources NLP*, que é o caso das línguas de sinais.

Nesse estudo, foi possível constatar através de uma série de experimentos que a adoção de uma dessas arquiteturas viáveis (*Transformer* básico ou ByT5) ajudaria a aumentar a precisão e qualidade do componente de tradução da Suíte VLibras, trazendo um aumento percentual máximo de até 12,73% nas traduções perfeitas, diminuindo as traduções incorretas em 16,21% e proporcionando uma melhoria média de 10,31% na métrica BLEU.

Os resultados demonstraram 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 e, até mesmo, para uma simplificação do componente tradutor da plataforma, tornando-o puramente neural.

Como continuação deste estudo, os autores planejam incluir outros conjuntos de dados na fase de avaliação da qualidade de tradução, incluindo frases e construções presentes em sites governamentais e privados, poesias, conteúdos literários e de outros contextos para melhorar a aferição do grau de generalização obtido pelos modelos avaliados. Outro possível trabalho futuro pode ser voltado para avaliar se os modelos *Transformer Básico* e ByT5 também podem proporcionar ganhos similares na qualidade de tradutores neurais usados em outras línguas de sinais.

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. (eds). **Emerging technologies in data mining and information security**. Lectures Notes in Networks and Systems, v. 164. Singapore: Springer, 2021. p. 359-369. 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*: 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*: 2019 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, p. 7784-7793, 2018. 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.; 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), 2021, Virtual. **Proceedings [...]**. 2021. Disponível em: https://iclr.cc/virtual/2021/poster/3013. Acesso em 02 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, Online. **Proceedings [...]**. 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: 2022 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, Online. **Proceedings [...]**. Online: ACL, 2021. 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, p. 117-124, 2019. 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.; SHEKTAR, R.; ALAM, M.; KAUR, R. Neural machine translation for low-resource languages: a survey. **ACM Computing Surveys**, v. 55, n. 11, p. 1-37, 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.. 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. (eds). **Computer Vision – ECCV 2020**. ECCV 2020. Lecture Notes in Computer Science(), v. 12356. Cham: Springer, 2020. 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, 2017. Disponível em https://openreview.net/pdf?id=B1ckMDqlg. Acesso em: 02 jan. 2024.

SOUZA, M. F. N. S.; ARAUJO, 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, 2005.

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, 2019. Disponível em: https://openreview.net/forum?id=SkVhlh09tX. Acesso em: 02 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, Virtual Event. **Proceedings [...]**. 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; INTERNATIONAL WORKSHOP ON AUTOMATIC TRANSLATION FOR SIGNED AND SPOKEN LANGUAGES (AT4SSL), 18., 1., 2021, Virtual USA. **Proceedings […].** p. 60-70. Disponível em: https://aclanthology.org/2021.mtsummit-at4ssl.7/. Acesso em: 22 dez. 2023.