

Uso de Glosas para Tradução Automática em Libras

Bruno Rogério Simões dos Santos
LAGIC, Universidade Estadual do Ceará
Fortaleza-CE, Brasil
bruno.rogerio@aluno.uece.br

Luan Fernandes de França
LADESC, Universidade Estadual do Ceará
Fortaleza-CE, Brasil
luan.franca@aluno.uece.br

Abstract—This article examines the challenge of communication between deaf and hearing people in Brazil, where more than 2 million people are hearing impaired and the proficiency in Libras is limited, exacerbated by the lack of interpreters. We propose developing a computational solution divided into two stages: first, converting the Portuguese text into an intermediate language, called gloss or interlingua, and then transforming this interlingua back into text through smoothing using a LLM. To validate the effectiveness of this approach, we compared its performance to an end-to-end direct translation model using the BLEU metric. The solution using gloss achieved a BLEU-1 of 71.23, while the direct model achieved a BLEU-1 of 34.01, representing a significant improvement in output quality. These results demonstrate that using gloss as an interlingua provides more faithful and natural translations than architectures that dispense with this intermediate step.

Keywords—Libras; Gloss; Translate.

Resumo—Este artigo examina o desafio da comunicação entre surdos e ouvintes no Brasil, onde mais de 2 milhões de pessoas têm deficiência auditiva e o domínio de Libras é restrito, agravado pela falta de intérpretes. Propõe-se desenvolver uma solução computacional dividida em duas etapas: primeiro, a conversão do texto em português para uma linguagem intermediária, denominada glosa ou interlíngua e, em seguida, a transformação dessa interlíngua em texto, novamente, através da suavização feita por uma LLM. Para validar a eficácia dessa abordagem, comparamos seu desempenho a um modelo fim a fim de tradução direta usando a métrica BLEU. A solução com uso de glosa alcançou um BLEU-1 de 71.23, enquanto o modelo direto obteve um BLEU-1 de 34.01, representando uma melhoria significativa na qualidade da saída. Esses resultados comprovam que o uso da glosa como interlíngua proporciona traduções mais fiéis e naturais do que arquiteturas que dispensam essa etapa intermediária.

Palavras-chave—Libras; Glosas; Tradução.

I. INTRODUÇÃO

De acordo com o Censo realizado pelo IBGE, em 2022, entre as 198,3 milhões de pessoas com dois anos ou mais de idade no país, 14,4 milhões (ou 7,3%) eram pessoas com deficiência. Deste número, 2,6 milhões de pessoas eram de

pessoas com deficiência auditiva.¹ A Organização Mundial da Saúde (OMS) estima que, até 2050, quase 2,5 bilhões de pessoas terão algum grau de perda auditiva, e mais de 700 milhões precisarão de reabilitação auditiva.²

Apesar do papel crucial da Língua Brasileira de Sinais (Libras) na redução de lacunas de comunicação, a proficiência geral permanece limitada, limitando a inclusão social e o nível de escolaridade. Pesquisas nacionais também revelam que apenas uma pequena fração dos surdos é totalmente proficiente em Libras, enquanto a escassez de intérpretes qualificados impede muitos de se envolverem plenamente na vida cotidiana e nas estruturas sociais.³

Para superar essas limitações, este trabalho propõe o emprego da glosa como interlíngua na cadeia de tradução automática entre Libras e português, que por meio de regras específicas cria-se um texto intermediário que pode ser convertido para a língua-alvo [1]. Ao adotar esse passo intermediário, o sistema pode explorar regras linguísticas e os mecanismos de atenção dos Transformers para gerar o texto em português de forma coerente e coesa. O artigo comparará o desempenho dessa abordagem com modelos fim a fim que realizam tradução direta das línguas de sinais para texto. A métrica BLEU será utilizada para avaliar a fidelidade e naturalidade das saídas. Demonstraremos que, ao inserir a glosa como facilitador, obtém-se ganhos significativos na qualidade da tradução, evidenciando a eficácia da interlíngua em relação às arquiteturas que dispensam essa representação intermediária.

II. FUNDAMENTAÇÃO TEÓRICA

A. Processamento de Linguagem Natural (PLN)

Processamento de Linguagem Natural (PLN) é uma sub-área da inteligência artificial que se concentra na interação

¹Disponível em: Agência de Notícias - IBGE. Acesso em: 11 de ago. de 2025.

²Disponível em: Organização Mundial da Saúde (OMS). Acesso em: 12 de ago. de 2025.

³Disponível em: PSN - Pesquisa Nacional de Saúde. Acesso em: 07 de jul. de 2025.

entre computadores e a linguagem humana. Seu objetivo principal é permitir que máquinas entendam, interpretem, gerem e respondam a textos ou falas de maneiras significativas e úteis. O PLN combina linguística computacional com aprendizado de máquina, aprendizado profundo e estatística para analisar e processar grandes volumes de dados em linguagem natural. Aplicações comuns incluem *chatbots*, tradução de idiomas, análise de sentimentos e assistentes de voz.

Uma etapa crítica no PLN é o pré-processamento de texto, que prepara dados brutos de linguagem para análise, removendo inconsistências e reduzindo a complexidade. Esse processo frequentemente envolve técnicas como tokenização, uso de letras minúsculas, remoção de pontuação e stopwords, e lematização ou *stemming*. Ao padronizar e estruturar a entrada, o pré-processamento melhora a qualidade e a eficiência dos modelos de PLN. Essas etapas são essenciais para converter dados não estruturados em formatos que podem ser usados por algoritmos de aprendizado de máquina para extrair padrões e fazer previsões [2].

Os avanços no PLN levaram ao desenvolvimento de modelos sofisticados como Word2Vec, GloVe, BERT e GPT, que conseguem capturar relacionamentos semânticos profundos e significado contextual na linguagem. Esses modelos melhoraram significativamente o desempenho dos sistemas de PLN em diversas tarefas, permitindo interações mais naturais e coerentes entre humanos e máquinas. À medida que a área continua a evoluir, espera-se que o PLN desempenhe um papel ainda maior em áreas como saúde, educação e atendimento ao cliente, transformando a forma como nos comunicamos e interagimos com a tecnologia.

B. Língua Brasileira de Sinais - LIBRAS

A Libras possui uma estrutura gramatical única, organizando-se por meio de parâmetros específicos que constituem os níveis linguísticos [3]. Similar às línguas faladas, a Libras tem a capacidade de gerar uma variedade infinita de mensagens ao combinar um conjunto limitado de elementos visuais, analogamente aos fonemas em línguas orais. Os elementos visuais na Libras são denominados parâmetros e incluem:

- **Configuração de mãos:** Refere-se à forma que a mão ou as mãos assumem ao realizar um sinal, sendo utilizado na produção de sinais datilológicos, *e.g.*, o alfabeto digital/manual. Este parâmetro é essencial para distinguir sinais que podem parecer similares, mas possuem configurações de mãos distintas;
- **Ponto de articulação ou locação:** Indica o local onde o sinal é articulado, podendo estar no corpo

do sinalizador (*e.g.*, na testa, no peito ou nos ombros) ou no espaço neutro, dependendo do sinal. Esse parâmetro é especialmente relevante em contextos onde há necessidade de clareza semântica, como em discursos mais formais;

- **Movimento:** Corresponde ao deslocamento que a mão realiza durante a execução do sinal. Pode ser executado por uma ou ambas as mãos, classificando-se como unidirecional, bidirecional ou multidirecional. Além disso, variações na intensidade ou repetição do movimento podem indicar aspectos gramaticais ou emocionais.;
- **Orientação da palma da mão:** Define a direção para a qual a palma da mão está voltada durante a realização do sinal, podendo influenciar diretamente o significado. Um mesmo movimento com orientações diferentes pode resultar em sinais com interpretações distintas;
- **Expressão facial e corporal:** Este parâmetro contribui para a transmissão da mensagem, permitindo, *e.g.*, a identificação de emoções ou se uma mensagem expressa exclamação ou interrogação. Por exemplo, uma expressão facial mais intensa pode sublinhar a gravidade ou a urgência da mensagem.

Além desses parâmetros, é importante mencionar que a Libras também apresenta aspectos sociolinguísticos próprios, como variações regionais e adaptações ao contexto cultural. Sinais podem variar entre regiões do Brasil, refletindo as influências locais e as experiências da comunidade surda em diferentes contextos. Ainda, o processo de aquisição da Libras, tanto como primeira língua para pessoas surdas quanto como segunda língua para ouvintes, envolve a compreensão desses parâmetros e de suas interações para uma comunicação eficaz [4].

C. Glosa - Interlíngua

A glosa não funciona como uma língua natural, tal qual o Português ou a Libras, mas sim como um sistema de notação que recorre a vocábulos de uma língua falada para representar, de modo aproximado, os sinais de uma língua de sinais. Seu propósito central é viabilizar a transcrição de dados de línguas de sinais para análises linguísticas, elaboração de dicionários ou desenvolvimento de soluções computacionais, como tradutores automáticos e avatares animados que sinalizam. Visto que as línguas de sinais ainda não dispõem de um sistema de escrita reconhecido de forma ampla, a glosa se apresenta como um recurso valioso para documentar e investigar esses idiomas [5].

As transcrições em glosa podem variar em nível de pormenorização. Na forma mais resumida, utiliza-se um conjunto básico de símbolos capazes de distinguir os sinais

essenciais, o que chamamos de transcrição “larga”. Já a transcrição “estreita” adiciona elementos que capturam detalhes mais finos do sinal, tais como configurações de mão ou movimentos específicos. Essa flexibilidade permite adaptar a profundidade da anotação conforme a finalidade do estudo ou aplicação tecnológica.

Entretanto, vale lembrar que a glosa é uma representação intermediária, dependente da língua falada de origem, o que implica limitações. Ela não consegue abarcar por completo a tridimensionalidade e simultaneidade inerentes às línguas de sinais, tornando desafiador reconstruir exatamente o sinal original apenas a partir da glosa [4]. Além disso, não existe um único padrão unificado de notação: diferentes projetos e pesquisadores podem adotar convenções distintas ou decidir em consenso qual glosa usar para representar determinados sinais.

D. Transformers

Transformers são um tipo de arquitetura de aprendizado profundo originalmente projetada para tarefas de processamento de linguagem natural, particularmente tradução automática. Ao contrário das redes neurais recorrentes (RNNs) tradicionais, os Transformers dependem de um mecanismo chamado autoatenção, que permite ao modelo considerar toda a sequência de entrada simultaneamente, identificando relações entre todos os elementos, independentemente de sua posição. Essa capacidade de processamento paralelo torna os Transformers altamente eficazes para a compreensão do contexto e da semântica em dados sequenciais, como linguagem [6].

Os modelos Transformer revolucionaram o processamento de sequências ao introduzir mecanismos de atenção que permitem ao sistema focar dinamicamente nas partes mais relevantes de um texto. No contexto da glosa, onde cada unidade lexical carrega informações sintéticas sobre movimento, configuração de mão e expressões faciais, os Transformers são capazes de captar dependências de longo alcance entre esses elementos, garantindo coesão e consistência na representação. Isso significa que, ao receber uma sequência em glosa, o Transformer identifica quais sinais devem ser priorizados ou suavizados para gerar um fluxo textual mais natural, de forma que seja possível ter um texto final coeso e coerente.

A LLM do DeepSeek, por exemplo, construída sobre a arquitetura Transformer de última geração, inclui um módulo de pós-processamento que atenua repetições desnecessárias e corrige saltos temáticos bruscos, gerando um texto mais fluido [7]. Ao combinar atenção contextual com probabilidades de sequência adaptadas, o sistema sugere alternativas de notação que respeitam tanto a morfologia das línguas de

sinais quanto a sintaxe do português. Com o uso de LLMs, em conjunto com uma linguagem intermediária (glosa) é possível obter resultados mais naturais e precisos do que abordagens diretas que não utilizam.

III. TRABALHOS RELACIONADOS

End-to-end mapping: O trabalho de [8] fez extração de features do I3D (Inflated 3D ConvNet) pré-treinado para tokenização de vídeos. A abordagem apresentada no trabalho é livre de glosas com tradução direta de vídeo para texto, sem representações intermediárias. Foi utilizado o *dataset How2Sign* com mais de 80 horas de vídeos ASL (*American Sign Language*) contínuos. Já os autores de [9] replicaram e melhoraram os resultados do trabalho anterior através de experimentos sistemáticos extensivos e otimização de hiperparâmetros. A implementação, independente da arquitetura original, foi testada com 36 configurações diferentes em uma série de experimentos sistemáticos. Essa replicação focou na otimização de funções de ativação, regularização e arquitetura e também fez uso das mesmas métricas (BLEU e rBLEU) para comparação direta.

A estrutura proposta no estudo de [10] introduziu o conceito de *Sign Language Translation* (SLT) como um campo distinto do reconhecimento de gestos tradicionais. Esse conceito vai além do reconhecimento ingênuo de gestos ao considerar as ricas estruturas gramaticais e linguísticas da língua de sinais para gerar traduções em linguagem falada a partir de vídeos. O trabalho estabelece que a tradução direta de vídeos de língua de sinais para texto falado é fundamentalmente diferente do reconhecimento sequencial de sinais individuais. Além disso, o trabalho ressalta que o uso de glosas melhora significativamente o desempenho da tradução de vídeos de línguas de sinais.

No trabalho [11] foi apresentado um conjunto de dados LIBRAS-UFOP multimodal de língua de sinais brasileira, de pares mínimos, usando um sensor Microsoft Kinect, desenvolvido com base no conceito de pares mínimos para reconhecimento de gestos isolados. O conjunto de dados é composto por 56 sinais divididos em 4 categorias, com 3.040 sequências de dados realizadas por 5 indivíduos, cada um executando aproximadamente 11 repetições por sinal. Os dados foram coletados por meio de um Microsoft Kinect V1 e incluem vídeos RGB-D, informações de profundidade e dados de esqueletos. O sistema de reconhecimento proposto utiliza as informações RGB-D, esqueleto e expressões verbais como entrada para três redes neurais profundas operando em paralelo. As características são extraídas dos vídeos por meio da técnica de *Rank Pooling*, que transforma sequências de frames em imagens dinâmicas representativas. As saídas dessas redes são combinadas por meio de

uma fusão baseada na média, resultando em uma única classificação final.

Mapping to glosses: O problema da segmentação temporal dos sinais em vídeo é determinante para o sucesso das abordagens do segundo tipo. Em [12], primeiramente é detectado pistas de sinais significativos ao longo dos vídeos e, então, aplica-se uma rede neural convolucional 3D com refinamento iterativo para resolver as ambiguidades entre pistas dos sinais detectados. O problema tratado em [13] é o de alinhar *gloss labels* ordenados aos sinais contidos em vídeo. Para identificar as fronteiras entre sinais, o problema é formulado como um problema de aprendizado por reforço. O estado é definido como sendo a representação de *features* de cada segmento de vídeo e a ação como sendo a determinação da localização da fronteira semântica entre símbolos. Já [14] apresenta um algoritmo baseado em assinaturas de atributos de diferença absoluta entre *frames* DAD (*Delayed Absolute Difference*) *signature* para detectar pausas e assim delimitar as fronteiras entre sinais em vídeos contínuos. A intenção de [15] é definir unidades significativas SU (*subtitle-unit*) como segmentos de vídeo correspondentes às fronteiras temporais de *glosas* que possam ser precisamente entendidas. Essas unidades são utilizadas para segmentar um SL (*Sign Language*) em sequências curtas e coerentes que possam ser traduzidas individualmente. Após essa segmentação, os autores aplicam uma rede neural convolucional baseada em grafos sobre as *features* para reconhecer as sentenças.

IV. METODOLOGIA

Nesta seção descrevemos em detalhes o fluxo desenvolvido para a geração automática de texto em glosa (representação da língua de sinais em forma textual) a partir de um texto em português, utilizando o framework Drools e o modelo de linguagem profunda (LLM) da OpenAI, GPT 4.o mini, bem como sua posterior “suavização” de volta para o português fluente por meio de outra LLM, que no caso deste trabalho foram o DeepSeek e o Gemma.

O Drools é um framework *open source* para sistemas baseados em regras (*Rule Engine*), desenvolvido em Java, muito usado para implementar lógica de negócios complexa de forma declarativa, separando regras do código imperativo tradicional. Ele não foi originalmente desenvolvido para processar linguagem de sinais ou linguagem glosa, mas pode ser adaptado para esse fim em projetos que envolvem: executar regras definidas sobre fatos linguísticos (tokens e categorias) ou armazenar as regras sobre sintaxe, estrutura de sentenças e tópicos, por exemplo. A linguagem usada para definir regras é a DRL (*Drools Rule Language*).

A. Aquisição e Pré-processamento

O texto-fonte em português é armazenado em um arquivo de texto, onde cada parágrafo é segmentado em sentenças usando expressões regulares que respeitam abreviações (Dr., Dra., Prof. etc.) para evitar quebras indevidas de frase. Em cada sentença isolada, remove-se toda pontuação gráfica (vírgulas, pontos, dois-pontos, ponto-e-vírgula, travessões), preservando apenas caracteres alfabéticos e espaços. Cada sentença “limpa” é então dividida em tokens por espaços em branco, gerando uma `List<String>` de unidades léxicas. Essa lista será o conjunto de entrada para o mecanismo de regras. A confecção das glosas foi baseada em trabalhos acadêmicos que já estabeleceram convenções e regras bem definidas para sua formulação. Ao usar o Drools como ambiente de execução de regras (*glosa.drl*) ou GPT 4.o mini, aplica-se em sequência as convenções listadas na Tabela I.

No Drools, por exemplo, após a execução de todas as regras (invocadas por `session.fireAllRules()` em uma *KieSession* isolada para cada sentença), coletamos a lista `s.getGlosa()` e concatenamos seus elementos em uma string única, separada por espaços. Esse texto livre de morfologia complexa e sem pontuação é então escrito em um arquivo de texto, por exemplo (`corpus_glosa.txt`), onde cada linha corresponde à glosa de uma sentença original. Já no GPT-4.o mini para gerar as glosas utilizamos um *prompt* que pedisse para a LLM fazer o mesmo pré-processamento realizado no Drools e aplicasse as mesmas convenções utilizadas para gerar as glosas das sentenças, ao final, foi passado um exemplo de entrada (sentença original em português) e um exemplo da resposta esperada (em glosa).

B. Suavização da Glosa via LLM

Como não dispomos de um *dataset* paralelo contendo sentenças em português e suas respectivas notações em glosa, adotamos um processo alternativo. Utilizamos o *dataset* *How2Sign* utilizado em [8] e [9] para aplicar nossa proposta e ao mesmo tempo validá-la. As sentenças em inglês do *dataset* foram traduzidas para o português e, em seguida, processadas por uma LLM (GPT 4.o mini) e pelo **Drools** para a geração das glosas correspondentes. Após essa etapa, aplicou-se a suavização das sentenças, simulando a tradução reversa das glosas para o português natural.

Os resultados são as sentenças suavizadas geradas a partir das glosas, utilizando técnicas de engenharia de *prompt* com a API do DeepSeek e Gemma. Após vários testes, o *prompt* utilizado nas suavizações foi: “Na tarefa de tradução de vídeo Libras para linguagem natural foi reconhecida a sequência de glosas dada. Gerar um texto

TABELA I
CONVENÇÕES PARA GERAÇÃO DA GLOSA.

Item	Convenção / Regra	Exemplo
Omissão de artigos	Omitir artigos definidos e indefinidos (o, a, os, as, um, uma etc.), pois Libras não possui esses marcadores.	CARRO RAPIDO
Flexão Verbal	Empregar verbos em forma base (sem flexão de pessoa/número). Indicar tempo por advérbios (ONTEM, JÁ, AGORA) posicionados no início da frase.	JOAO JA COMER
Pronomes Possessivos	Sujeitos e pronomes pessoais geralmente omitidos ou indicados por apontamento (IX) quando claros no contexto. Para possessivos, usar POSS(referente) conforme convenção.	CASA POSS(ELE)
Negação	Inserir NÃO após o verbo principal para negação simples. Para o verbo TER no sentido “não ter”, mantém-se a convenção TER- NÃO .	ELE NAO TER
Adjetivos e Advérbios	Adjetivos vêm após o substantivo qualificado, sem flexão de gênero/número. Advérbios de modo/intensidade seguem o verbo.	CASA GRANDE
Marcadores Temporal ou Espacial	Colocar advérbios temporais (ONTEM, HOJE, AMANHA) no início da frase para marcar o tempo. Marcadores espaciais (AQUI, LA, CIDADE etc.) ou apontamentos para local também podem vir antes da oração.	ONTEM MARIA JOAO VER
Omissão de Preposições e Conjunções	Omitir preposições e conjunções, pois esses conectivos estão incorporados ao próprio sinal da Libras ou inferidos pelo contexto.	EU IR CASA

suave correspondente. Sequência: xxx. A escolha inicial dessas LLMs foi por conta da gratuidade do uso dessas APIs, mesmo com um limite de requisições diárias.

C. Métrica de Avaliação

O BLEU (Bilingual Evaluation Understudy) é uma métrica amplamente utilizada na avaliação automática de sistemas de tradução automática [16]. Proposto para reduzir a dependência de avaliações humanas, o BLEU mede a similaridade entre uma sentença candidata e um conjunto de traduções de referência, com base na correspondência de n-gramas. Sua fórmula combina a média geométrica das precisões n-gram modificadas, ponderadas por pesos w_n , com um fator de penalidade de brevidade (Brevity Penalty — BP), que penaliza traduções muito curtas. A métrica é expressa como:

$$\text{BLEU} = \text{BP} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right),$$

onde p_n representa a precisão n-gram modificada para n-gramas de ordem n . A penalidade de brevidade é calculada com base na razão entre o comprimento total das traduções candidatas (c) e o comprimento efetivo das referências (r), definida como $\text{BP} = 1$ se $c > r$ e $\text{BP} = e^{(1-r/c)}$ caso contrário. O BLEU varia de 0 a 1, sendo frequentemente escalado para a faixa de 0 a 100.

Em resumo, o BLEU avalia o quanto uma frase gerada por um modelo se aproxima de uma frase de referência, verificando a sobreposição de *n-gramas*, ou seja, de sequências de palavras. Na prática, o BLEU-1 mede o quanto as palavras individuais coincidem (avalia vocabulário e escolha lexical), o BLEU-2 analisa pares de palavras (avaliando pequenas combinações e contexto local), enquanto o BLEU-3 e o BLEU-4 consideram trechos mais longos, refletindo a fluidez, a coesão e a ordem das frases. Assim, valores mais altos de BLEU-4 indicam que a sentença é mais natural e próxima da estrutura esperada de uma glosa.

V. RESULTADOS

A Tabela II apresenta os resultados da avaliação automática das sentenças suavizadas geradas a partir das glosas geradas pelo Drools, utilizando técnicas de engenharia de prompt com a API do DeepSeek. Observa-se que as suavizações obtiveram pontuações satisfatórias nas métricas BLEU, com destaque para valores mais altos no BLEU-1 (representado na tabela como BLEU), que avalia a correspondência de 1-grama entre as sentenças de referência e as suavizadas. Isso indica uma boa sobreposição lexical e estrutural, apesar das reformulações naturais de estilo e ordem.

Fornecemos também uma avaliação qualitativa dos resultados na Tabela II, destacando algumas das sentenças em

português, suas versões em glosa e as suavizações geradas. O exemplo (01) demonstra a capacidade do modelo em produzir traduções detalhadas e precisas, com pontuações elevadas em BLEU e rBLEU, indicando uma tradução fluente e semanticamente coerente. Por outro lado, o exemplo (02) revela as limitações, com quedas acentuadas nas métricas, sugerindo dificuldades do modelo ao lidar com frases mais longas ou com estruturas mais complexas. Os resultados sugerem que o uso da notação em glosa como linguagem intermediária é uma abordagem promissora para suavização de textos em Libras.

TABELA II
EXEMPLOS DE AVALIAÇÃO COM rBLEU E BLEU (DROOLS / DEEPSEEK)

Exemplo	Tipo	Texto	rBLEU	BLEU
(01)	PT-BR GLOSA	Você não precisa dar passos grandes. VOCE PRECISAR DAR PASSOS GRANDES NAO	100.00	100.00
	GS	Você não precisa dar passos grandes.		
(02)	PT-BR	Ele é movido pelos pés e é o que realmente direciona o avião para a direita e esquerda quando você inicia e para uma curva.	64.92	64.92
	GLOSA	ELE MOVIDO PELOS PES REALMENTE DIRECIONAR AVIAO DIREITA ESQUERDA QUANDO VOCÊ INICIAR CURVA		
	GS	Ele é controlado pelos pés e é o que realmente direciona o avião para a direita ou para a esquerda quando você inicia uma curva.		

Já na Tabela III, a glosa foi gerada pela *GPT-4.o mini* e suavizada pela *Gemma*, que é uma LLM do Google. Para fins de comparação, utilizamos as mesmas sentenças e a mesma métrica. Um exemplo em que a suavização feita pelo DeepSeek apresentou melhor resultado foi quando a ferramenta manteve o tom descritivo da sentença original, ao contrário da Gemma, que transformou a sentença em uma pergunta (Exemplo (01)), alterando completamente o sentido. O mesmo ocorreu no Exemplo (02), que envolvia uma sentença mais longa. A suavização realizada pelo DeepSeek demonstrou melhor desempenho em ambos os casos, tanto em sentenças curtas quanto sentenças longas.

Analisando os trabalhos mais recentes publicados na área de tradução automática de Libras, observa-se uma tendência crescente no uso de modelos baseados em *deep learning* com abordagem de aprendizado fim-a-fim para tradução contínua (*Continuous SLR*). Esses modelos têm demonstrado desempenho promissor, superando os métodos tradicionais baseados em segmentação seguida de reconhecimento de sinais isolados. No entanto, ainda enfrentam desafios importantes, como a necessidade de grandes volumes de dados anotados para o treinamento eficaz e a dificuldade de implementação em dispositivos de borda, devido ao

TABELA III
EXEMPLOS DE AVALIAÇÃO COM rBLEU E BLEU (GPT 4.0 MINI / GEMMA)

Exemplo	Tipo	Texto	rBLEU	BLEU
(01)	PT-BR GLOSA	Você não precisa dar passos grandes. VOCÊ PRECISAR DAR PASSO GRANDE NAO	6.03	6.03
	GS	Você precisa dar um passo grande, não?.		
(02)	PT-BR	Ele é movido pelos pés e é o que realmente direciona o avião para a direita e esquerda quando você inicia e para uma curva.	16.15	16.15
	GLOSA	HOJE ELE MOVIDO PÉ POSS(ELE) DIRECIONAR AVIAO DIREITA ES- QUERDA QUANDO VOCÊ INICIAR CURVA NAO		
	GS	Hoje, ele está se movendo com o pé na direção do avião, virando para a direita e para a esquerda quando você iniciar uma curva. Não.		

alto consumo de memória e processamento requerido por modelos de grande porte.

VI. DISCUSSÃO

Os resultados comparativos em relação à métrica BLEU estão apresentados na Tabela IV, onde se pode observar que as abordagens propostas superaram os métodos fim-a-fim nas métricas rBLEU e BLEU em todos os níveis (BLEU-1 a BLEU-4), evidenciando seu potencial como alternativa eficiente e precisa na tradução automática de Libras. Tanto a abordagem utilizando o *DeepSeek* quanto a que utilizou o *Gemma* para a suavização das glosas mostraram ter mais clareza semântica nas sentenças quando comparadas aos modelos TLS (Tradução de Linguagem de Sinais) fim-a-fim dos trabalhos de [8] e [9].

Uma comparação entre as construções das glosas com o *Drools* e a suavização feita com o *DeepSeek* e as glosas geradas pelo *GPT-4.o mini* e suavizadas com o *Gemma* mostrou que os resultados da primeira abordagem foram superiores. As sentenças finais geradas pela combinação *Drools + DeepSeek* apresentaram valores mais expressivos nas métricas BLEU, indicando uma melhor correspondência com as glosas de referência. Mesmo utilizando o mesmo *prompt* para as duas abordagens, as saídas geradas pelo *Drools* mostraram maior clareza, fidelidade e consistência linguística, o que pode ter influenciado diretamente para um melhor resultado alcançado pelo *DeepSeek* em comparação ao *Gemma*. Esses resultados comparativos são apresentados na Tabela V.

Em relação à glosa não conseguir abarcar por completo a tridimensionalidade e simultaneidade inerentes às línguas de sinais, a arquitetura Transformer, por meio do mecanismo de atenção, permite relacionar qualquer parte de

TABELA IV
COMPARAÇÃO DAS MÉTRICAS BLEU ENTRE OS TRABALHOS AVALIADOS

Abordagens	rBLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Modelo TLS fim-a-fim [8]	2.21	34.01	19.30	12.18	8.03
Modelo TLS fim-a-fim [9]	2.43	33.86	19.26	12.12	7.94
Drools + DeepSeek	41.47	71.23	59.09	49.54	41.47
GPT 4.o mini + Gemma	22.05	52.40	37.76	28.71	22.05

TABELA V
COMPARAÇÃO DAS MÉTRICAS BLEU (SENTENÇA COM GLOSA | SENTENÇA GLOSA SUAVIZADA)

Abordagens	rBLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Sentença + Drools	24.71	57.91	43.55	32.74	24.71
Drools + DeepSeek	41.47	71.23	59.09	49.54	41.47
Sentença + GPT 4.o mini	11.04	43.91	27.21	16.92	11.04
GPT 4.o mini + Gemma	22.05	52.40	37.76	28.71	22.05

entrada com qualquer outra. Isso teoricamente facilita modelar eventos simultâneos porque um *token* pode "prestar atenção" em vários sinais ou canais ao mesmo tempo. Logo, para contornar essa limitação da glosa, o uso de Transformers será fundamental no suporte para implementação de um modelo avatar para tradução bidirecional levando em conta os gestos da língua de sinais (pose, mãos e rosto).

VII. CENÁRIOS DE USO

A solução proposta pode ser aplicada em diversos contextos onde a comunicação entre ouvintes e usuários de Libras é essencial. Em serviços públicos e de atendimento ao cidadão, por exemplo, o uso da glosa como interlíngua pode ser integrado a *chatbots* e centrais de atendimento automatizadas, permitindo que sistemas de resposta em português processem inicialmente a glosa extraída de vídeos em Libras e gerem respostas claras e naturais em português. Essa integração pode ser realizada por meio de APIs de processamento de linguagem natural (como o *OpenAI API* ou o *Hugging Face Transformers*) conectadas a plataformas de atendimento, como o *Dialogflow* ou o *Microsoft Bot Framework*.

Além disso, em plataformas de mídia social e de produção de conteúdo online, criadores de vídeos em Libras podem empregar esta abordagem para gerar automaticamente legendas em português, utilizando o modelo de glosa e a suavização feita por LLMs. Essa automação amplia o alcance comunicativo de conteúdos em Libras e favorece a acessibilidade digital.

No contexto de avatares 3D e agentes virtuais, a solução proposta pode atuar como um agente de IA de decisão linguística, responsável por selecionar a melhor forma de

gerar a glosa com base em um banco limitado de animações disponíveis para o avatar. Diferentemente de abordagens literais, como a adotada no *V-Libras*, este agente considera não apenas a tradução palavra a palavra, mas também a adequação semântica e contextual da frase. Assim, o modelo pode decidir qual estrutura de glosa é mais apropriada para ser representada de forma fluida e natural na animação do sinal, equilibrando fidelidade linguística e viabilidade de execução.

Por fim, em contextos de pesquisa linguística e desenvolvimento tecnológico, o corpus paralelo (glosa → português ou português → glosa) produzido pelo método pode servir como base para o treinamento e avaliação de novos modelos de tradução e síntese de linguagem de sinais, fomentando avanços contínuos na área.

VIII. AMEAÇAS À VALIDADE

Uma das principais ameaças à validade interna reside na precisão das regras Drools para a geração de glosa, pois depende fortemente da cobertura das convenções linguísticas estabelecidas. Omissões ou conflitos entre regras podem introduzir vieses na representação interlíngua, afetando a posterior suavização pelo LLM. Quanto à validade externa, a generalização do método para outras línguas de sinais ou variantes regionais de Libras pode ser limitada, uma vez que o conjunto de regras e os exemplos de treinamento do DeepSeek foram calibrados para o português brasileiro e a Libras padrão. Diferenças dialetais em configurações de mão, expressões faciais e ordem sintática podem comprometer a eficácia da solução em comunidades surdas específicas. Por fim, a validade de construção pode ser ameaçada por aspectos inerentes aos LLMs, como tendência

a “alucinações” e repetição de padrões aprendidos que não correspondem ao sentido original. Embora o módulo de pós-processamento mitigue parte desses problemas, ele não elimina completamente erros de coerência e possíveis distorções semânticas.

IX. CONCLUSÕES

Este trabalho teve o intuito de mostrar a eficácia do uso de glosas na tradução entre a língua de sinais e a língua escrita/falada. As abordagens apresentadas utilizando glosas mostraram-se ser superior ao modelo de tradução direta fim-a-fim, com ganhos expressivos na métrica BLEU que é uma métrica que avalia a qualidade das traduções. Como trabalhos futuros, pretende-se montar um dataset em Libras, contínuo, a partir de fontes abertas como TV Senado, TV Cultura ou YouTube, implementar e avaliar um algoritmo reconhecimento de língua de sinais em vídeo contínuo, implementar e avaliar um gerador de prompt para modelos LLMs pré-treinados (GPT ou DeepSeek) para gerar textos suaves compreensíveis a partir de uma sequência de glosas de sinais detectada, implementar um modelo de avatar para tradução bidirecional e realizar uma avaliação comparativa de desempenho com outros trabalhos publicados (estado da arte).

AGRADECIMENTOS

Agradecemos especialmente aos professores doutores José Everardo Bessa Maia e Leonardo Sampaio Rocha pela orientação e pelos comentários críticos durante o desenvolvimento deste trabalho.

REFERÊNCIAS

- [1] C. Ellwanger, C. Santos, P. Soares, and T. Weisheimer, “G’libras: Um sistema especialista para auxílio à criação de glosas,” in *Workshop de Desafios da Computação aplicada à Educação (DesafIE!)*. SBC, 2015, pp. 101–110.
- [2] J. Eisenstein, “Natural language processing,” *Jacob Eisenstein*, vol. 507, 2018.
- [3] C. L. Ferraz, *Dicionário de Configuração de Mãos em Libras*. Cruz das Almas, Brasil: Editora UFRB, 2019.
- [4] R. M. Quadros, *Língua de herança: língua brasileira de sinais*. Porto Alegre, RS: Penso Editora, 2017.
- [5] F. A. dos Santos Paiva, J. M. De Martino, P. A. Barbosa, Á. Benetti, and I. R. Silva, “Um sistema de transcrição para língua de sinais brasileira: o caso de um avatar,” *Revista do GEL*, vol. 13, no. 3, pp. 12–48, 2016.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [7] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi *et al.*, “Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning,” *arXiv preprint arXiv:2501.12948*, 2025.
- [8] L. Tarrés, G. I. Gállego, A. Duarte, J. Torres, and X. Giró-i Nieto, “Sign language translation from instructional videos,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5625–5635.
- [9] P. Roy, J.-E. Han, S. Chouhan, and B. Thumu, “American sign language video to text translation,” *arXiv preprint arXiv:2402.07255*, 2024.
- [10] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, “Neural sign language translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [11] L. R. Cerna, E. E. Cardenas, D. G. Miranda, D. Menotti, and G. Camara-Chavez, “A multimodal libras-ufop brazilian sign language dataset of minimal pairs using a microsoft kinect sensor,” *Expert Systems with Applications*, vol. 167, p. 114179, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420309143>
- [12] K. Renz, N. C. Stache, S. Albanie, and G. Varol, “Sign language segmentation with temporal convolutional networks,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 2135–2139.
- [13] C. Wei, J. Zhao, W. Zhou, and H. Li, “Semantic boundary detection with reinforcement learning for continuous sign language recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 1138–1149, 2020.
- [14] S. Khan, D. G. Bailey, and G. S. Gupta, “Pause detection in continuous sign language,” *International journal of computer applications in technology*, vol. 50, no. 1-2, pp. 75–83, 2014.
- [15] H. Bull, M. Gouiffès, and A. Braffort, “Automatic segmentation of sign language into subtitle-units,” in *European Conference on Computer Vision*. Springer, 2020, pp. 186–198.
- [16] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th Annual Meeting of the ACL*. Philadelphia, PA, USA: Association for Computational Linguistics, 2002, pp. 311–318.