De Sinais a Texto: Um Sistema Inteligente para Tradução Automática de Libras com Foco na Inclusão Educacional

Ana Mara de Oliveira Figueiredo, Aline Vilela Guarisi, Camila Feres Valinho, Gabriely Vicente S Marcelino, Vivia Mery Souza,

[1]Instituto Federal Fluminense Campos Bom Jesus do Itabapoana

{ana.figueiredo, camila.valinho, vivia.souza}@iff.edu.br

{alinevguarisi, gabient.ut}@gmail.com

Abstract. *The Brazilian Sign Language (Libras) is essential for communication within the deaf community but still faces barriers in translation to written Portuguese. The lack of effective tools limits accessibility and interaction between deaf and hearing individuals. This work proposes a system based on artificial intelligence and image processing to convert Libras signs into text. The development involved video collection, preprocessing, and machine learning model training. Preliminary results indicate potential for accurate and automated translation. The solution aims to promote social and educational inclusion and assist hearing individuals interested in learning Libras.*

Resumo. *A Língua Brasileira de Sinais (Libras) é fundamental para a comunicação da comunidade surda, mas ainda enfrenta barreiras na tradução para o português escrito. A ausência de ferramentas eficazes limita a acessibilidade e a interação entre surdos e ouvintes. Este trabalho propõe um sistema baseado em inteligência artificial e processamento de imagens para converter sinais de Libras em texto. O desenvolvimento envolveu coleta e pré-processamento de vídeos, além do treinamento de modelos de aprendizado de máquina. Os resultados preliminares indicam potencial para uma tradução precisa e automatizada. A solução visa promover inclusão social e educacional, além de auxiliar ouvintes interessados em aprender Libras.*

1. Introdução

A Língua Brasileira de Sinais (Libras) é uma língua natural de modalidade gestual-visual, estruturada com gramática própria e essencial para a comunicação e expressão cultural da comunidade surda no Brasil. Seu reconhecimento oficial ocorreu com a sanção da Lei nº 10.436, em 2002 [1], representando um marco importante para a valorização da identidade surda e para a promoção de sua inclusão social [Brasil].

Segundo dados do Censo Demográfico de 2022, divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), aproximadamente 2,3 milhões de brasileiros possuem alguma deficiência auditiva [Instituto Brasileiro de Geografia e Estatística 2022]. Embora uma parte significativa dessa população utilize a Libras como principal meio de comunicação,a saber, 22,4%, ainda existem diversas barreiras que dificultam sua plena

[1]Utiliza-se neste estudo, portanto, a nomenclatura seguindo o padrão legislativo, sendo assim grafado e nomeado Libras (Língua Brasileira de Sinais).

DOI: 10.5753/sbie.2025.12170

inclusão social, tais como a escassez de intérpretes qualificados e a baixa familiaridade da sociedade em geral com a língua de sinais [Martins 2016].

Historicamente, a ausência de políticas públicas adequadas e de reconhecimento oficial da Libras levou, durante décadas, à adoção de práticas educacionais oralistas, que priorizavam a fala e a leitura labial em detrimento da língua de sinais. Como apontado por [de Freitas Reis and de Morais 2021], essa abordagem gerou isolamento social e prejuízos emocionais para muitos surdos, que frequentemente criavam sistemas gestuais próprios restritos ao ambiente familiar.

A regulamentação da Libras, bem como a formação de intérpretes especializados, trouxe avanços significativos no acesso das pessoas surdas aos sistemas educacional, jurídico e de saúde. Contudo, persistem desafios relacionados à comunicação com ouvintes e ao acesso à informação em espaços onde não há intérpretes disponíveis [Martins 2016].

Neste contexto, o desenvolvimento de tecnologias assistivas baseadas em inteligência artificial (IA) surge como uma estratégia promissora para minimizar as barreiras comunicacionais enfrentadas pela comunidade surda. A automatização da tradução de sinais da Libras para o Português escrito pode facilitar o acesso à informação e, ao mesmo tempo, atuar como ferramenta auxiliar no processo de ensino-aprendizagem, tanto para pessoas surdas quanto para ouvintes que desejam aprender Libras.

Este trabalho propõe o desenvolvimento de um sistema baseado em técnicas de visão computacional e aprendizado profundo, capaz de reconhecer sinais da Libras e traduzí-los automaticamente para o Português escrito. É importante destacar que, nesta fase inicial, o sistema realiza o reconhecimento exclusivamente de sinais estáticos, como aqueles utilizados para representar letras e alguns sinais isolados. A expansão para o reconhecimento de sinais dinâmicos — característicos de grande parte das construções frasais na Libras — constitui um desafio adicional, atualmente em andamento.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta os principais trabalhos relacionados na área de tradução automática de sinais. A Seção 3 descreve detalhadamente a metodologia adotada no desenvolvimento do sistema. A Seção 4 apresenta e analisa os resultados obtidos. Por fim, a Seção 5 discute as conclusões e propõe direções para trabalhos futuros.

2. Trabalhos Relacionados

A tradução automática de Libras tem sido explorada por diferentes abordagens, variando entre soluções baseadas em sensores, visão computacional e uso de redes neurais profundas. Feliciano et al. [Feliciano et al. 2023] utilizaram redes LSTM combinadas com camadas densas para o reconhecimento de sinais estáticos e dinâmicos em ambientes complexos, alcançando resultados relevantes. Entretanto, diferentemente dessa proposta, este estudo adota uma arquitetura híbrida CNN-LSTM, que possibilita capturar tanto características visuais (formas, bordas, texturas) quanto temporais, equilibrando precisão e eficiência para aplicações em tempo real.

No campo educacional, Machado e Pinheiro [Machado and Pinheiro 2022] ressaltam a importância de ferramentas que facilitem o ensino de Libras para ouvintes, incentivando a interação entre surdos e não-surdos. Nesse sentido, o sistema aqui proposto

diferencia-se por priorizar a tradução de Libras para o português escrito, contribuindo não apenas para acessibilidade, mas também como ferramenta de apoio pedagógico.

Outros trabalhos exploram sensores de movimento, como Tavares et al. [Tavares et al. 2010], que capturam sinais por meio de dispositivos adicionais. Embora eficazes em alguns contextos, tais soluções limitam a aplicabilidade cotidiana. Em contraste, nosso modelo fundamenta-se exclusivamente em visão computacional, dispensando sensores externos e privilegiando praticidade e escalabilidade.

Entre soluções já consolidadas, destacam-se o aplicativo *Hand Talk* [Hand Talk], e o software LIBROL [Carvalho et al. 2013], ambos voltados para a tradução do português para Libras. Também merece menção o projeto Rybená [Moreira et al. 2011], que converte textos em sinais animados por avatar. Contudo, essas iniciativas trabalham majoritariamente na direção português $\rightarrow$ Libras. Nossa proposta segue caminho inverso — Libras $\rightarrow$ português — ainda pouco explorado, mas de alta relevância para ampliar a comunicação inclusiva.

Rocha et al. [Rocha et al. 2020] propuseram uma abordagem baseada em CNN para tradução de Libras em tempo real, obtendo 90% de acurácia. Nosso trabalho avança nesse sentido ao integrar LSTM à arquitetura convolucional, permitindo lidar melhor com dependências temporais de sinais dinâmicos e viabilizando geração textual gramaticalmente correta por meio de técnicas de PLN.

Por fim, Banerjee et al. [Banerjee et al. 2022] apresentam uma revisão abrangente sobre reconhecimento de línguas de sinais com IA, destacando desafios como diversidade de sinais, contexto e expressões faciais. Essa análise reforça as escolhas metodológicas aqui adotadas, voltadas a construir um sistema escalável, educacionalmente útil e aplicável em diferentes condições de uso.

Em síntese, enquanto estudos anteriores avançam em reconhecimento ou tradução unidirecional, nosso trabalho diferencia-se por combinar CNN-LSTM e PLN em um sistema orientado à tradução direta de Libras para o português escrito, com foco em acessibilidade educacional.

3. Metodologia

A metodologia deste trabalho baseou-se em [Feliciano et al. 2023], que utilizou redes LSTM com camadas densas para o reconhecimento de sinais estáticos e dinâmicos da Língua Brasileira de Sinais (Libras), Contudo, optou-se por modificar a arquitetura, implementando um modelo híbrido CNN-LSTM, capaz de capturar de forma eficiente tanto as características visuais (espaciais) quanto temporais, proporcionando maior robustez e eficiência no reconhecimento em tempo real.

A Figura 1 ilustra o fluxograma aplicado neste trabalho. O passo 1 envolve a coleta dos dados, na qual foi construído uma base de dados contendo vídeos de letras estáticas e sinais dinâmicos, utilizando o dataset V-LIBRASIL e um dataset próprio. No passo 2, ilustra-se a realização do pré-processamento dos vídeos, que consiste na conversão dos vídeos em sequências de frames, normalização e identificação da região de interesse (ROI) para otimizar a extração de características visuais. No passo 3 demonstra-se a implementação e o treinamento do modelo CNN-LSTM para mapear sinais para palavras correspondentes em Português. Em sequência, o passo 4 refere-se ao reconhecimento de

sinais, onde o modelo treinado foi aplicado em novos vídeos, realizando a predição das palavras associadas às configurações manuais em tempo real. No passo 5, foi utilizado um dicionário e técnicas de PLN para gerar frases completas e gramaticalmente corretas. Por fim, no passo 6, conclui-se com o desenvolvimento de uma interface de visualização, permitindo que o usuário veja o vídeo original, a tradução em texto e as legendas geradas. A interface foi desenvolvida como uma aplicação web a partir do Flask, tornando-o acessível a partir de qualquer computador com um navegador moderno e uma webcam conectada.

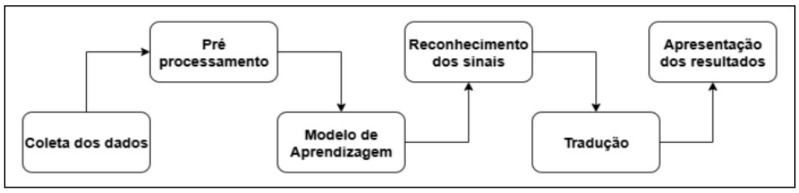

Figura 1. Fluxograma da metodologia utilizada neste trabalho

3.1. Coleta e pré-processamento de dados

Os dados utilizados foram provenientes do *dataset* V-LIBRASIL e de um *dataset* próprio. O primeiro forneceu uma ampla gama de sinais dinâmicos sob variadas condições de iluminação e ângulos, enquanto o segundo consistiu em aproximadamente 300 imagens por letra do alfabeto da Libras, capturadas com a *Webcam* garantindo diversidade na qualidade e nos ângulos das imagens.

Os vídeos e imagens foram processados para gerar sequências de *frames*, organizadas em diretórios conforme as classes de sinais. Para a extração de pontos de referência das mãos, utilizou-se o *MediaPipe Hands*, configurado no modo de imagem estática, garantindo a identificação precisa de 21 pontos de referência por mão. As coordenadas extraídas foram normalizadas e organizadas em duas listas: uma contendo os dados processados e outra associando cada conjunto à sua respectiva classe. Posteriormente, os dados foram serializados e armazenados no arquivo *data.pickle*, facilitando o carregamento para o treinamento do modelo.

A conversão de vídeos em *frames* foi realizada com a biblioteca *OpenCV*, possibilitando a transformação de vídeos em sequências de imagens que servem como entrada para a rede neural. Cada *frame* foi redimensionado para 224x224 *pixels*, assegurando padronização e eficiência na extração de características visuais. A detecção da região de interesse (ROI) também foi aplicada, delimitando um retângulo central no frame e reduzindo ruídos de fundo, melhorando a qualidade dos dados para o treinamento.

Esses procedimentos garantiram a consistência e a qualidade necessárias para o treinamento do modelo, enquanto o gerenciamento eficiente dos arquivos e o uso de paralelismo foram essenciais para otimizar o processamento de grandes volumes de dados.

3.2. Treinamento do modelo

O modelo foi desenvolvido com uma combinação de redes convolucionais (CNN) e redes LSTM. A CNN foi composta por duas camadas convolucionais, a primeira com 16 filtros de tamanho 3x3 e a segunda com 32 filtros 3x3, ambas seguidas por camadas

de *max pooling* 2x2 e ativação ReLU. Essas camadas extraíram características espaciais (bordas, formas e texturas) dos sinais. A saída da CNN foi então passada a uma LSTM, responsável por capturar dependências temporais entre os *frames*, fundamentais para o reconhecimento de sinais dinâmicos.

O treinamento foi realizado utilizando tensores tridimensionais compostos pelas sequências de *frames*. Os dados foram divididos em 80% para treinamento e 20% para validação, organizados em *batches* de 32 amostras. A biblioteca *PyTorch* foi utilizada para implementar o modelo e gerenciar automaticamente os gradientes, além de possibilitar a aceleração com o uso de GPUs.

A função de perda utilizada foi a *CrossEntropyLoss*, adequada para tarefas de classificação, pois mede a diferença entre as probabilidades previstas e as classes reais. O otimizador escolhido foi o Adam, que ajusta dinamicamente a taxa de aprendizado e acelera a convergência.

O treinamento ocorreu por múltiplas épocas, com monitoramento contínuo da função de perda e da acurácia em conjunto de validação. Além disso, foram utilizadas a matriz de confusão para análise dos erros de classificação e a evolução da perda (*loss curve*) para identificar situações de subajuste (*underfitting*) ou sobreajuste (*overfitting*). Esses indicadores auxiliaram no ajuste do número de épocas e dos hiperparâmetros do modelo, garantindo maior consistência dos resultados.

3.3. Reconhecimento de sinais

O processo de reconhecimento contemplou tanto sinais estáticos (letras e palavras isoladas) quanto sinais dinâmicos (palavras e frases curtas). Para ambos os casos, as etapas seguiram um *pipeline* baseado em visão computacional, que envolveu a captura de imagens/vídeos, extração de características, normalização e predição do sinal correspondente.

No caso dos sinais estáticos, foram utilizadas imagens capturadas por *webcam* em diferentes ângulos. A extração de características foi realizada com o *MediaPipe Hands*, que detectou 21 pontos bidimensionais (x, y) das articulações das mãos. As coordenadas foram normalizadas e armazenadas em vetores de características, posteriormente utilizados para treinar classificadores implementados com a biblioteca *Scikit-learn*.

Para os sinais dinâmicos, o reconhecimento foi conduzido pelo modelo CNN-LSTM descrito anteriormente. As sequências de frames processadas pela CNN forneceram características espaciais, enquanto a LSTM modelou as dependências temporais entre os frames, permitindo capturar os movimentos necessários à interpretação correta dos sinais.

Esse fluxo integrado permitiu o reconhecimento em tempo real tanto de sinais estáticos quanto dinâmicos, assegurando maior abrangência do sistema e aproximando-o de cenários de uso reais.

4. Resultados

Os primeiros testes realizados tiveram como objetivo validar a capacidade do sistema em traduzir sinais de Libras para o português escrito. Para isso, os experimentos foram conduzidos em duas etapas complementares. A primeira etapa envolveu testes práticos com voluntários, nos quais foram coletadas amostras de sinais realizados em tempo real.

A segunda etapa consistiu na análise quantitativa do desempenho do modelo por meio de métricas consolidadas em tarefas de reconhecimento de padrões.

Os testes iniciais foram realizados com um grupo fechado de cinco usuários, cujas mãos apresentavam diferentes tamanhos e tonalidades de pele, para avaliar o desempenho do algoritmo em diversas condições. Cada usuário realizou sinais referentes a cada uma das letras do alfabeto três vezes, totalizando 315 amostras. Devido à representação dos sinais muito semelhante, o algoritmo apresentou dificuldades em diferenciar as letras A, E e S, resultando em detecções imprecisas. O próximo passo será expandir os testes para um grupo maior de usuários, permitindo uma avaliação mais robusta do modelo, além de ajustes no algoritmo conforme as melhorias identificadas.

Outro exemplo é apresentado na Figura 2 exemplifica o funcionamento do sistema ao reconhecer o sinal da letra “C”. O caractere previsto é exibido na tela em tempo real, mostrando a tradução automática do sinal. Pode-se observar que o sistema manteve a eficácia sob diferentes condições de iluminação, posicionamento e distância da câmera, demonstrando a capacidade de generalização do modelo interpretação do sinal.

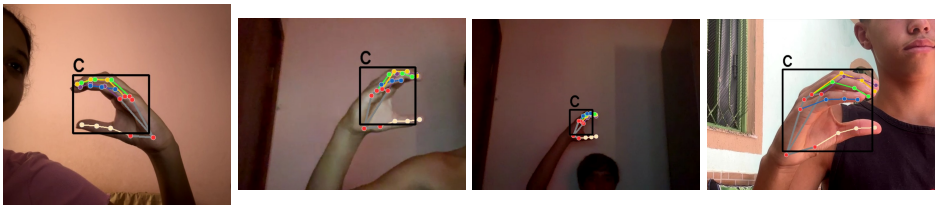

Figura 2. Exemplos de reconhecimento da letra ’C’

Após os testes com sinais estáticos, a segunda etapa do projeto concentrou-se no treinamento do sistema para reconhecer sinais dinâmicos, iniciando com expressões simples, .como “Obrigado” e “De nada”, ilustradas na Figura 3, aproximando o funcionamento do sistema de um contexto mais acessível ao usuário final. A captura do sinal exigiu a interação manual para registrar os sinais, o que resultou em um tempo de processamento ligeiramente maior em relação aos sinais estáticos, embora ainda inferior a um minuto por execução. Apesar dessa diferença, o sistema demonstrou resultados satisfatórios, conseguindo interpretar corretamente os gestos dinâmicos e apresentar a tradução em português escrito na interface.

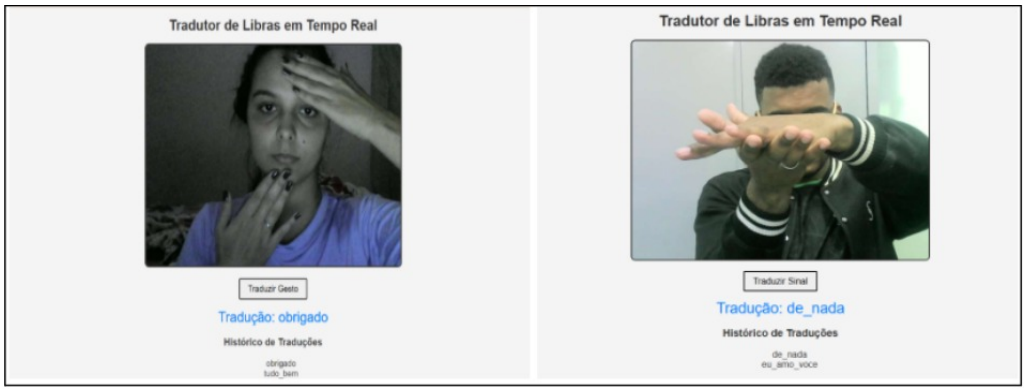

Figura 3. Testes com a Interface Funcional.

Em termos quantitativos, os experimentos mostraram que, devido à limitação do *dataset*, foi necessário reduzir o vocabulário para seis palavras e limitar o treinamento a 15 épocas. Essa estratégia buscou equilibrar a complexidade do modelo com o volume de dados disponível, evitando problemas de sobreajuste. Como resultado, o sistema alcançou uma acurácia de aproximadamente 84%, superando os primeiros testes e apresentando maior capacidade de generalização.

A análise da matriz de confusão referente ao melhor modelo obtido evidenciou que a maioria dos sinais foi corretamente reconhecida, com destaque para classes como "comi_muito" e "eu_amo_voce", que apresentaram taxas de acerto próximas ou superiores a 90%. Por outro lado, as confusões mais recorrentes ocorreram entre sinais com configurações manuais semelhantes, reforçando a necessidade de ampliar a base de dados para melhor discriminar essas classes.

De forma complementar, os testes realizados com voluntários em ambiente de webcam demonstraram que o sistema é capaz de traduzir tanto sinais estáticos quanto dinâmicos em português escrito de forma satisfatória. Embora os sinais dinâmicos tenham demandado um tempo ligeiramente maior de processamento na interface web, o desempenho global foi consistente, indicando que os ajustes realizados no treinamento contribuíram para maior robustez do modelo. Essas observações reforçam o potencial da proposta como ferramenta acessível e aplicável em contextos educacionais inclusivos.

5. Conclusão

O desenvolvimento do sistema de tradução de Libras para o português representa uma contribuição relevante para a promoção da acessibilidade e da inclusão social, com especial impacto no contexto educacional. Ao facilitar a comunicação entre surdos e ouvintes, o sistema pode atuar como um recurso pedagógico valioso em salas de aula inclusivas, ampliando as possibilidades de aprendizagem para estudantes surdos e contribuindo para a formação de uma comunidade escolar mais sensível à diversidade linguística. Além disso, pode ser utilizado por professores e alunos ouvintes como ferramenta didática para o ensino e a aprendizagem da Libras de forma prática e acessível.

Os resultados iniciais do projeto, especialmente no reconhecimento de sinais estáticos como o alfabeto em Libras, mostram-se promissores e indicam o potencial da tecnologia no apoio à mediação didática e na superação de barreiras linguísticas. Métricas como acurácia, precisão e recall reforçam a viabilidade do modelo em ambientes reais, embora ainda existam desafios a serem enfrentados, como a variação de sinais entre usuários, condições de iluminação e a necessidade de respostas em tempo real. A etapa de pós-processamento, com uso de técnicas de Processamento de Linguagem Natural (PLN), revelou-se essencial para garantir uma tradução mais coerente e adequada ao contexto educacional.

A próxima fase do projeto busca ampliar a base de dados para abranger sinais dinâmicos e desenvolver um protótipo funcional aplicável em diferentes contextos de ensino e aprendizagem. A capacidade futura de traduzir frases completas e expressões compostas permitirá interações mais naturais, o que pode beneficiar tanto o ensino bilíngue quanto práticas pedagógicas inclusivas. Assim, este trabalho reforça a importância da aplicação da inteligência artificial na educação como meio de reduzir desigualdades, valorizar a Libras como língua de instrução e construir caminhos mais efetivos para a inclusão

de estudantes surdos no processo educacional.

Referências

Banerjee, K., Harsha K, G., Kumar, P., Vats, I., P, V., Dasila, P., Akhtar, N., Kumar, A., and Gautam, P. (2022). A review on artificial intelligence based sign language recognition techniques.

Brasil. Dispõe sobre a língua brasileira de sinais - libras e dá outras providências.

Carvalho, R. S., Brito, J. O., Rodrigues, J. P., Silva, I. Q., Matos, P. F., and Oliveira, C. R. S. d. (2013). Librol: Software tradutor de português para libras. In *Encontro Nacional de Computação dos Institutos Federais (ENCOMPIF)*, pages 2118–2121, Maceió. Sociedade Brasileira de Computação.

de Freitas Reis, M. B. and de Morais, I. C. V. (2021). Inclusão dos surdos no brasil: do oralismo ao bilinguismo. *Revista UFG*, 20(1):1–20. Disponível online.

Feliciano, F. D. d. O., Briano, A. d. A., Prudencio, R. B. C., and Alves, E. C. (2023). Recognition of static or dynamic libras words in complex background environments. In *Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–4, Aveiro, Portugal. IEEE.

Hand Talk. Sobre o hand talk – acessibilidade em libras. Acesso em: 20 jul. 2024.

Instituto Brasileiro de Geografia e Estatística (2022). Censo demográfico 2022: População e domicílios - primeiros resultados.

Machado, S. M. d. S. and Pinheiro, R. C. (2022). O ensino de libras para estudantes ouvintes como um meio de inclusão de surdos. *Revista Panorâmica*, 36.

Martins, V. R. O. (2016). Educação bilíngue de surdos e diferenças: diálogo ainda necessário? SciELO em Perspectiva: Humanas. Acesso em: 22 de Julho de 2024.

Moreira, J. R., Ferneda, E., Brito, P. H., Coradine, L. C., Guadagnin, R. d. V., Oliveira, R. M. d., and Garcia, E. d. V. (2011). Rumo a um sistema de tradução português-libras. In *Workshop de Informática na Escola (WIE)*, pages 1543–1552, Aracajú. Sociedade Brasileira de Computação.

Rocha, J., Lensk, J., Ferreira, T., and Ferreira, M. (2020). Towards a tool to translate brazilian sign language (libras) to brazilian portuguese and improve communication with deaf. In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–4, Dunedin, New Zealand.

Tavares, J. E., Barbosa, J. L., and Leithardt, V. R. (2010). Sensorlibras: Tradução automática libras-português através da computação ubíqua. In *II Congresso Nacional de Pesquisa em Tradução e Interpretação de Língua de Sinais Brasileira*, pages 1–7.