Revista Ibero-Americana de Humanidades, Ciências e Educação — REASE

doi.org/10.51891/rease.v12i3.25311

ARQUITETURA HÍBRIDA PARA TRADUÇÃO AUTOMÁTICA DE PORTUGUÊS BRASILEIRO PARA LIBRAS EM CONTEXTO EDUCACIONAL: UMA ABORDAGEM BASEADA EM SEQ2SEQ, DICIONÁRIO DE MAPEAMENTOS E WORD EMBEDDINGS

Sarah Vitória Cantanhede Andrade[1]
Juliany Pereira Costa[2]
Thiago Nelson Faria dos Reis[3]

RESUMO: Este artigo apresenta a arquitetura híbrida desenvolvida para um sistema de tradução automática de português brasileiro para glossa da Língua Brasileira de Sinais (Libras), voltado ao suporte de alunos surdos em salas de aula regulares. A abordagem combina três estratégias complementares de tradução: (i) um modelo neural Sequence-to-Sequence (Seq2Seq) com redes Long Short-Term Memory (LSTM), treinado sobre corpus proprietário de 800 pares de sentenças; (ii) um dicionário de mapeamentos diretos com aproximadamente 500 entradas de alta precisão; e (iii) word embeddings treinados com a biblioteca Gensim para tratamento de tokens fora do vocabulário (OOV). A integração é gerenciada por uma camada de orquestração baseada no padrão Strategy, viabilizando a seleção dinâmica da melhor estratégia para cada sentença. O sistema incorpora ainda reconhecimento automático de fala offline (motor Vosk) e comunicação assíncrona via WebSocket, atingindo latência total de 150–350 ms e suporte a 10 sessões simultâneas. Os resultados indicam acurácia de 95% para sentenças presentes no corpus de treinamento e 33% para sentenças inéditas, com confiança de tradução entre 80 e 95%. O trabalho valida a viabilidade técnica da abordagem e aponta a expansão do corpus e a adoção de arquiteturas Transformer como direções prioritárias de evolução.

Palavras-chave: Tradução Automática. Libras. Arquitetura Híbrida. Seq2Seq. Processamento de Linguagem Natural.

ABSTRACT: This paper presents the hybrid architecture developed for an automatic translation system from Brazilian Portuguese to Brazilian Sign Language (Libras) gloss, designed to support deaf students in mainstream classrooms. The approach combines three complementary translation strategies: (i) a Sequence-to-Sequence (Seq2Seq) neural model with Long Short-Term Memory (LSTM) networks, trained on a proprietary corpus of 800 sentence pairs; (ii) a high-precision direct mapping dictionary with approximately 500 entries; and (iii) word embeddings trained with the Gensim library for out-of-vocabulary (OOV) token handling. Module integration is managed by an orchestration layer based on the Strategy design pattern, enabling dynamic selection of the best strategy per sentence. The system also incorporates offline automatic speech recognition (Vosk engine) and asynchronous WebSocket communication, achieving total latency of 150–350 ms and supporting 10 simultaneous sessions. Results indicate 95% accuracy for training corpus sentences and 33% for unseen sentences, with translation confidence between 80 and 95%. The work validates the technical feasibility of the hybrid approach and identifies corpus expansion and Transformer-based architectures as priority future directions.

Keywords: Automatic Translation. Libras. Hybrid Architecture. Seq2Seq. Natural Language Processing.

[1]Graduanda de Sistemas de Informação, Centro Universitário Santa Terezinha (CEST).
[2]Mestranda em Ciência da Computação - Universidade Federal do Maranhão (UFMA). Graduada pela Faculdade Santa Terezinha (CEST).
[3]Orientador: Doutor, Pós-Doutorando, Universidade Federal do Maranhão (UFMA).

1 INTRODUÇÃO

A inclusão educacional de alunos com deficiência auditiva representa um desafio persistente no sistema de ensino brasileiro. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2022), aproximadamente 5% da população brasileira possui algum grau de deficiência auditiva, totalizando mais de 10 milhões de pessoas, das quais 2,7 milhões apresentam surdez profunda. Para essa parcela da população, a Língua Brasileira de Sinais (Libras) constitui o principal meio de comunicação, sendo reconhecida oficialmente pela Lei n.º 10.436/2002 e regulamentada pelo Decreto n.º 5.626/2005, que estabelece a obrigatoriedade de intérpretes nas instituições de ensino.

Apesar dos avanços normativos, a demanda por intérpretes de Libras supera largamente a oferta de profissionais habilitados, especialmente em municípios do interior do país. Ferramentas de tradução automática disponíveis ainda apresentam limitações consideráveis, resultando em perda de informação durante a tradução e comprometendo o aprendizado de alunos surdos (CORRÊA et al., 2017). Essa lacuna é especialmente crítica na educação básica, onde a comunicação em tempo real é fundamental para o acompanhamento das aulas.

O campo do Processamento de Linguagem Natural (PLN) e do Aprendizado de Máquina tem avançado expressivamente, abrindo novas possibilidades para sistemas de tradução automática. Arquiteturas neurais como as redes Sequence-to-Sequence (Seq2Seq) com Long Short-Term Memory (LSTM) demonstraram resultados promissores em tarefas de tradução entre línguas naturais (SUTSKEVER; VINYALS; LE, 2014), incluindo línguas de sinais (DA ROSA ZUCOLOTTO et al., 2019). Mais recentemente, a introdução dos mecanismos de atenção (BAHDANAU; CHO; BENGIO, 2015) e da arquitetura Transformer (VASWANI et al., 2017) impulsionou o estado da arte em tradução automática.

Paralelamente, o reconhecimento automático de fala (ASR) beneficiou-se de avanços significativos com soluções de código aberto como Vosk e Whisper, tornando essa tecnologia acessível a desenvolvedores sem infraestrutura de grande porte (ALPHACEPHEI, 2020; RADFORD et al., 2023). A combinação de ASR com modelos de tradução neural cria a possibilidade de pipelines ponta a ponta capazes de converter fala em língua de sinais de forma automática e em tempo quase real.

Diante desse cenário, o presente artigo propõe e descreve em profundidade a arquitetura híbrida de um sistema integrado que captura a fala do professor em tempo real, converte-a em texto por reconhecimento automático de fala e a traduz para representações textuais em glossa

de Libras utilizando três estratégias complementares de tradução. O objetivo central é detalhar as decisões arquiteturais que permitem combinar robustez, baixa latência e operação offline, tornando a solução viável para implantação em escolas públicas brasileiras.

2 REFERENCIAL TEÓRICO

2.1 TRADUÇÃO AUTOMÁTICA PARA LÍNGUAS DE SINAIS

A tradução automática para línguas de sinais enfrenta desafios específicos em relação à tradução entre línguas orais escritas, pois envolve a conversão entre modalidades linguísticas distintas: oral-auditiva e visoespacial. A maioria dos sistemas existentes adota a glossa como representação intermediária, uma notação textual que representa sinais por palavras em maiúsculas, permitindo que modelos de PLN operem em espaço discreto antes de uma eventual animação de avatares (STOLL et al., 2018).

O projeto VLibras, amplamente utilizado em contextos institucionais brasileiros, representa uma solução baseada em dicionário e regras gramaticais. Embora funcional para frases de alta frequência, não disponibiliza métricas formais de avaliação de qualidade de tradução, dificultando comparações objetivas (SILVA et al., 2021). Trabalhos acadêmicos como Freitas et al. (2019) reportam cobertura de aproximadamente 70% do vocabulário de domínio educacional com abordagens baseadas em regras.

Para sistemas neurais, o corpus PHOENIX-2014T, utilizado por Camgoz et al. (2018) para tradução em Língua de Sinais Alemã, contém aproximadamente 7.096 pares de sentenças de treinamento. Modelos Seq2Seq treinados nesse corpus atingiram BLEU-4 de 9,58, enquanto modelos Transformer posteriores alcançaram BLEU-4 de 21,80 (CAMGOZ et al., 2020), demonstrando o potencial das arquiteturas baseadas em atenção.

2.2 ARQUITETURA SEQ2SEQ COM LSTM

A rede Long Short-Term Memory (LSTM), proposta por Hochreiter e Schmidhuber (1997), é um tipo especial de rede neural recorrente projetada para superar o problema do desvanecimento do gradiente, que impede que redes recorrentes simples aprendam dependências de longo prazo. A LSTM introduz um estado de célula que percorre toda a sequência com modificações controladas por três portas: de esquecimento (forget gate), de entrada (input gate) e de saída (output gate).

A arquitetura Sequence-to-Sequence (Seq2Seq), introduzida por Sutskever, Vinyals e Le (2014), utiliza um encoder LSTM para condensar a sequência de entrada em um vetor de contexto, e um decoder LSTM para gerar iterativamente a sequência de saída. Essa arquitetura demonstrou excelente desempenho em tarefas de tradução de máquina neural (NMT), sendo particularmente adequada para pares linguísticos com diferenças estruturais significativas, como o português e a glossa de Libras.

2.3 WORD EMBEDDINGS E PADRÕES DE DESIGN

Word embeddings são representações vetoriais densas de palavras em espaços semânticos contínuos, treinados de forma não supervisionada sobre grandes corpora (MIKOLOV et al., 2013). O algoritmo Skip-gram, implementado na biblioteca Gensim (Word2Vec), aprende a prever palavras do contexto a partir de uma palavra central, produzindo vetores que capturam relações semânticas por proximidade geométrica.

A adoção de padrões de design consolidados na engenharia de software orientada a objetos é fundamental para a manutenibilidade e extensibilidade de sistemas de PLN em produção. O padrão Strategy permite encapsular algoritmos intercambiáveis em objetos separados, tornando possível substituir a estratégia de tradução sem modificar o código cliente (GAMMA et al., 1994). O padrão Observer habilita a arquitetura orientada a eventos necessária para comunicação assíncrona em tempo real.

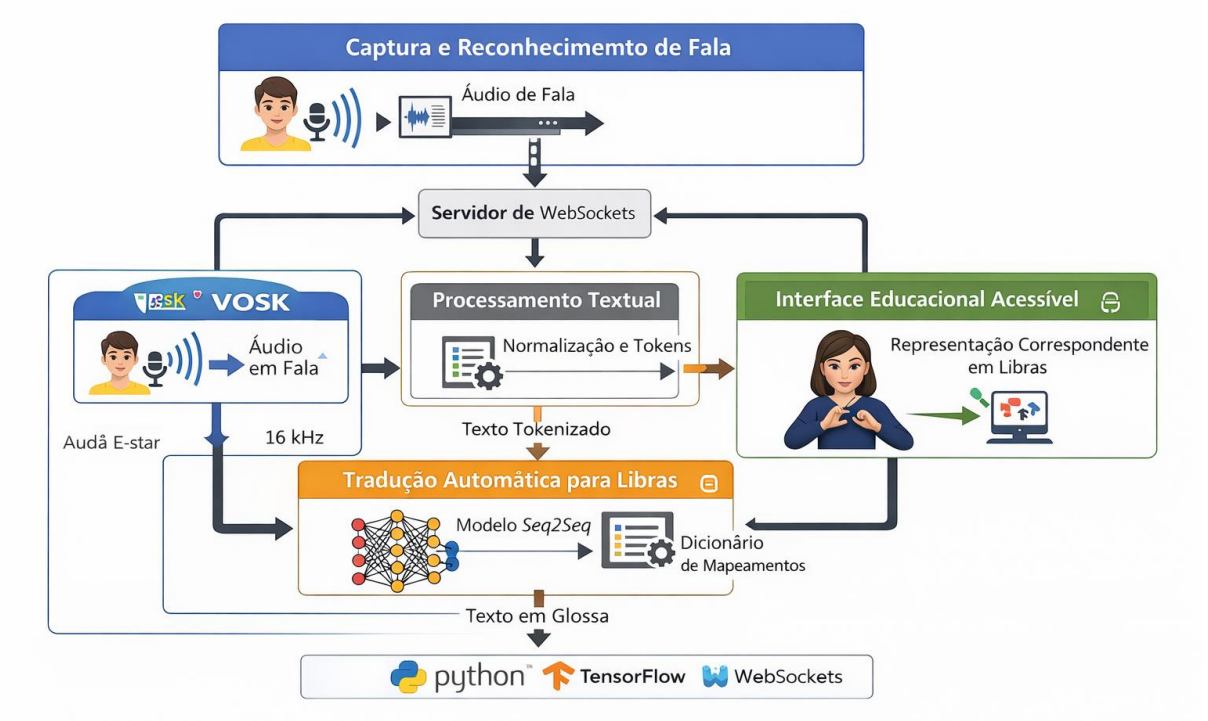
3 METODOLOGIA E ARQUITETURA DO SISTEMA

3.1 VISÃO GERAL DA ARQUITETURA HÍBRIDA

O sistema foi projetado segundo uma arquitetura de microsserviços orientada a serviços, com cinco componentes principais e comunicação assíncrona entre si, adotando os princípios SOLID e os padrões arquiteturais de microsserviços para favorecer a manutenibilidade, testabilidade e escalabilidade independente de cada componente.

Os cinco módulos são: (1) Módulo de Captura e Reconhecimento Automático de Fala (ASR); (2) Módulo de Pré-processamento Textual; (3) Módulo de Tradução Automática para Libras; (4) Servidor de Comunicação em Tempo Real (WebSocket); e (5) Interface Educacional Acessível. A arquitetura foi guiada por quatro princípios fundamentais: Separação de Responsabilidades, Baixo Acoplamento, Rastreabilidade (logs estruturados com Session IDs e timestamps ISO-8601) e Escalabilidade.

Figura 1 - Fluxo de dados do pipeline fala-texto-glossa. As setas indicam o sentido do fluxo de informação entre os módulos. A comunicação entre o módulo ASR e o módulo de tradução é assíncrona, mediada pelo servidor WebSocket.



Foram adotados cinco padrões de design: Singleton (carregamento do modelo ASR em memória única), Factory (criação de objetos de transcrição), Strategy (seleção dinâmica da estratégia de tradução), Observer (gerenciamento de eventos WebSocket) e Dependency Injection (configuração centralizada dos componentes). Esses padrões garantem qualidade arquitetural e aderência às boas práticas de desenvolvimento orientado a objetos.

3.2 MÓDULO DE RECONHECIMENTO AUTOMÁTICO DE FALA

O reconhecimento automático de fala foi implementado com o motor Vosk (ALPHACEPHEI, 2020), solução de código aberto baseada na biblioteca Kaldi que oferece modelos pré-treinados para português brasileiro e opera completamente offline. A escolha pelo Vosk em detrimento de alternativas baseadas em nuvem foi motivada por: (i) independência de conectividade à internet; (ii) ausência de custos recorrentes por uso de API; e (iii) controle total sobre os dados de áudio dos usuários, aspecto relevante para a privacidade em ambientes escolares.

O módulo opera com taxa de amostragem de 16 kHz, canal mono e formato PCM de 16 bits. O processamento ocorre em modo de análise contínua de frames de áudio, gerando transcrições parciais em tempo quase real. Um mecanismo de controle de qualidade filtra transcrições com base em comprimento textual mínimo e limiar de confiança configurável, reduzindo a propagação de erros de reconhecimento para as etapas subsequentes do pipeline. A acurácia observada para português brasileiro situou-se entre 85% e 90% em condições controladas.

3.3 CORPUS LINGUÍSTICO E PRÉ-PROCESSAMENTO

O corpus utilizado para treinamento do modelo neural é um recurso proprietário composto por 800 pares paralelos de sentenças em português brasileiro e suas representações em glossa de Libras, com foco no domínio educacional. A padronização terminológica das glossas seguiu as convenções do Instituto Nacional de Educação de Surdos (INES): sinais em letras maiúsculas, verbos na forma infinitiva e pronomes pessoais por extenso (EU, VOCÊ, ELE/ELA).

A Tabela 1 apresenta exemplos representativos dos mapeamentos presentes no corpus.

Tabela 1 – Exemplos de pares paralelos do corpus português–Libras

Sentença em Português	Glossa em Libras
Bom dia, como você está?	BOM DIA VOCÊ COMO ESTAR
Eu quero aprender Libras.	EU QUERER APRENDER LIBRAS
Abra o livro na página dez.	ABRIR LIVRO PÁGINA DEZ
Ele precisa de ajuda.	ELE PRECISAR AJUDA
A professora está explicando.	PROFESSORA EXPLICAR
Você entendeu a matéria?	VOCÊ ENTENDER MATÉRIA

O pipeline de pré-processamento incluiu: normalização ortográfica (correção de erros de digitação, padronização de acentuação e caixa tipográfica), codificação inteira de tokens, inserção de tokens especiais <start> e <end> nas sequências de saída, e padding pós-sequencial para uniformização do comprimento das sequências. A divisão dos dados seguiu a proporção 80/20: 640 pares para treinamento e 160 para validação.

3.4 AS TRÊS ESTRATÉGIAS DE TRADUÇÃO E A CAMADA DE ORQUESTRAÇÃO

A principal contribuição arquitetural do sistema é a combinação de três estratégias complementares de tradução, gerenciadas por uma camada de orquestração baseada no padrão Strategy. Para cada sentença de entrada, o orquestrador determina dinamicamente qual estratégia aplicar, maximizando cobertura e precisão.

A primeira estratégia é o dicionário de mapeamentos diretos, serializado em formato pickle com aproximadamente 500 entradas organizadas em categorias funcionais: cumprimentos e despedidas (~30), pronomes pessoais (~15), verbos de alta frequência (~120), adjetivos básicos (~80), expressões pedagógicas (~50) e demais categorias (~205). O dicionário é consultado prioritariamente quando todos os tokens da sentença possuem mapeamento direto, garantindo máxima precisão e velocidade para o vocabulário de alta frequência. A integração com o modelo Seq2Seq é gerenciada pelo padrão Strategy: se todos os tokens possuírem cobertura, o dicionário é utilizado diretamente; caso contrário, a sentença é encaminhada ao modelo neural.

A segunda estratégia é o modelo neural Seq2Seq com LSTM, ativado para sentenças fora da cobertura total do dicionário. A arquitetura é composta por uma camada de Embedding para o idioma de entrada, um Encoder LSTM, um Decoder LSTM e uma camada densa com ativação softmax. O treinamento foi conduzido por 500 épocas com otimizador Adam (lr = 0,001), função de perda Categorical Cross-Entropy, batch size de 64 e modo teacher forcing, atingindo loss de 3,7978e-05 e val_loss de 1,2925e-05.

A terceira estratégia são os word embeddings, utilizados como mecanismo de fallback para tokens fora do vocabulário (OOV). Um modelo Word2Vec foi treinado com a biblioteca Gensim sobre as 800 glossas do corpus proprietário, com vetores de dimensão 100, algoritmo Skip-gram e janela de contexto de 5 tokens. Para tokens desconhecidos, o sistema busca o vizinho mais próximo no espaço de embeddings, expandindo a cobertura léxica sem necessidade de retreinamento.

O pipeline completo de processamento percorre as seguintes etapas em ordem: (i) captura do sinal de áudio pelo microfone do professor (16 kHz, mono, PCM 16-bit); (ii) processamento incremental de frames pelo motor Vosk com estimativa de confiança; (iii) verificação de critérios de qualidade; (iv) pré-processamento textual; (v) seleção da estratégia de tradução pelo orquestrador; (vi) execução da tradução; (vii) pós-processamento com fallback

OOV via word embeddings; e (viii) transmissão da glossa gerada para a interface educacional via WebSocket, com rastreamento por Session ID e timestamp ISO-8601.

4 RESULTADOS E DISCUSSÃO

4.1 CONVERGÊNCIA DO MODELO NEURAL

O modelo Seq2Seq apresentou convergência estável ao longo das 500 épocas de treinamento, com redução progressiva e consistente das métricas loss e val_loss, e desaceleração gradual da taxa de queda a partir de aproximadamente 300 épocas. Ao final do treinamento, o modelo atingiu loss de treinamento de 3,7978e-05 e val_loss de 1,2925e-05. O fato de o val_loss ter atingido valor inferior ao loss de treinamento pode ser explicado pela composição do conjunto de validação, que pode incluir desproporcionalmente sentenças mais curtas ou mais frequentes no corpus.

4.2 DESEMPENHO OPERACIONAL DO PIPELINE

As métricas de desempenho operacional foram mensuradas em ambiente local controlado, com 30 amostras por categoria de sentença. A Tabela 2 consolida os principais resultados obtidos.

Tabela 2 – *Métricas de desempenho operacional do pipeline completo*

Métrica	Valor Observado	Condições de Medição
Latência STT (Vosk)	100–300 ms	Sentenças curtas a médias; ambiente silencioso
Latência de tradução (Seq2Seq)	10–50 ms	CPU local; sentenças de 2–6 palavras
Latência total (fala → glossa)	150–350 ms	Pipeline completo; ambiente controlado
Throughput máximo	10 sessões simultâneas	Ambiente local; degradação controlada acima desse limite
Acurácia STT (português BR)	85–90%	Domínio educacional; locutor único; microfone externo
Confiança de tradução	80–95%	Estimativa probabilística; varia com complexidade da sentença

O tempo médio de resposta total inferior a 350 ms é relevante para a aplicação pretendida: estudos de percepção humana indicam que atrasos abaixo de 500 ms são geralmente imperceptíveis em contextos de comunicação mediada por tecnologia (NIELSEN, 1993). O suporte a até 10 sessões simultâneas é suficiente para cobrir uma sala de aula típica com múltiplos dispositivos de alunos conectados ao mesmo servidor local.

4.3 AVALIAÇÃO QUALITATIVA DAS TRADUÇÕES

A avaliação qualitativa foi conduzida com 50 sentenças distribuídas em três categorias. A Tabela 3 sumariza os resultados.

Tabela 3 – Resultados da avaliação qualitativa das traduções

Categoria de Sentença	Qtd.	Acerto (%)	Principais Erros Observados
Sentenças do conjunto de treinamento	20	95%	Inversões de ordem em sentenças complexas
Sentenças estruturalmente reorganizadas	15	67%	Omissão de adjuntos adverbiais
Sentenças completamente inéditas	15	33%	Traduções incompletas; perda de contexto; tokens OOV

Os resultados confirmam o padrão esperado para modelos Seq2Seq treinados com corpora de pequena escala: excelente desempenho em sentenças vistas durante o treinamento, com queda progressiva à medida que aumenta a distância estrutural em relação ao corpus. Para sentenças completamente inéditas, a taxa de acerto de 33% demonstra que o modelo capturou alguns padrões generalizáveis, sendo insuficiente para uso prático sem a complementação do dicionário de mapeamentos diretos.

4.4 LIMITAÇÕES E COMPARAÇÃO COM A LITERATURA

A principal limitação observada é a dificuldade de generalização para sentenças inéditas, característica de fenômenos de overfitting em modelos treinados com corpora de pequena escala. Para contextualizar, o corpus PHOENIX-2014T, utilizado por Camgoz et al. (2018) para tradução em Língua de Sinais Alemã, contém aproximadamente 7.096 pares de sentenças — cerca de 9 vezes o tamanho do corpus utilizado neste trabalho. A comparação direta com trabalhos para Libras especificamente é dificultada pela escassez de corpora públicos anotados para essa língua.

A segunda limitação significativa é a perda de contexto em sequências longas, uma limitação arquitetural intrínseca ao Seq2Seq tradicional, que comprime toda a sequência de entrada em um único vetor de contexto de dimensão fixa. A solução estabelecida são os mecanismos de atenção (BAHDANAU; CHO; BENGIO, 2015), que permitem ao decoder acessar diretamente todos os estados ocultos do encoder. A implementação de Bahdanau

attention sobre a arquitetura LSTM atual constitui um próximo passo natural antes da migração completa para Transformers.

Em contrapartida, a abordagem proposta apresenta diferenciais relevantes: (i) operação completamente offline do motor Vosk, crítica para escolas públicas com conectividade instável; (ii) latência total de 150–350 ms comparável à de soluções em nuvem com boa conectividade; (iii) arquitetura modular com o padrão Strategy, que permite a integração transparente de novos modelos sem alteração dos demais módulos; e (iv) abordagem híbrida com três camadas de tradução que maximiza cobertura e robustez frente às limitações individuais de cada componente.

5 CONSIDERAÇÕES FINAIS

Este artigo apresentou e detalhou a arquitetura híbrida de um sistema de tradução automática de português brasileiro para glossa de Libras, combinando três estratégias complementares: modelo neural Seq2Seq com LSTM, dicionário de mapeamentos diretos de alta precisão e word embeddings como fallback para tokens OOV. A abordagem foi validada como prova de conceito, demonstrando viabilidade técnica para implantação em ambientes educacionais com infraestrutura limitada.

Os resultados obtidos evidenciam que a combinação de estratégias híbridas maximiza a cobertura e a robustez do sistema frente às limitações individuais de cada componente. A latência total abaixo de 350 ms e o suporte a 10 sessões simultâneas validam a adequação do sistema para o contexto de sala de aula. O trabalho contribui para o avanço das investigações sobre soluções computacionais voltadas à inclusão e à acessibilidade linguística.

Como direções prioritárias para trabalhos futuros, destacam-se: (i) expansão do corpus de treinamento para pelo menos 5.000 pares de sentenças; (ii) implementação de mecanismos de atenção (Bahdanau attention) sobre a arquitetura LSTM atual; (iii) migração para arquiteturas baseadas em Transformers; (iv) integração com sistemas de animação de avatares para representação gestual completa de Libras; e (v) validação com usuários surdos em contexto escolar real, incluindo avaliação de usabilidade e eficácia pedagógica.

REFERÊNCIAS

ALPHACEPHEI. Vosk: Offline Speech Recognition API. alphacephei.com, 2020. Disponível em: https://alphacephei.com/vosk. Acesso em: 10 set. 2024.

BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. Neural machine translation by jointly learning to align and translate. In: International Conference on Learning Representations. San Diego: ICLR, 2015. p. 1–15.

CAMGOZ, Necati Cihan et al. Neural sign language translation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018. p. 7784–7793.

CAMGOZ, Necati Cihan et al. Sign language transformers: joint end-to-end sign language recognition and translation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020. p. 10023–10033.

CORRÊA, Ana F. et al. Desafios na interpretação de Libras em contextos educacionais: uma análise qualitativa. Revista Brasileira de Educação Especial, Marília, v. 23, n. 2, p. 211–226, abr./jun. 2017.

DA ROSA ZUCOLOTTO, Deivis et al. Tecnologias assistivas para surdos: revisão sistemática da literatura. Cadernos de Educação, Pelotas, n. 62, p. 1–22, 2019.

FREITAS, Jéssica A. et al. Sistema de tradução automática português-Libras baseado em dicionário e regras gramaticais para o domínio educacional. Revista de Informática Teórica e Aplicada, Porto Alegre, v. 26, n. 3, p. 45–60, 2019.

GAMMA, Erich et al. Design Patterns: Elements of Reusable Object-Oriented Software. Reading: Addison-Wesley, 1994.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. Neural Computation, Cambridge, v. 9, n. 8, p. 1735–1780, nov. 1997.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Demográfico 2022: pessoas com deficiência. ibge.gov.br, 2022. Disponível em: https://www.ibge.gov.br. Acesso em: 10 set. 2024.

MIKOLOV, Tomas et al. Efficient estimation of word representations in vector space. In: International Conference on Learning Representations. Scottsdale: ICLR, 2013. p. 1–12.

NIELSEN, Jakob. Usability Engineering. San Francisco: Morgan Kaufmann, 1993.

RADFORD, Alec et al. Robust speech recognition via large-scale weak supervision. In: International Conference on Machine Learning. PMLR, 2023. p. 1–20.

SILVA, Ricardo et al. VLibras: uma solução de acessibilidade digital em Libras para o governo federal brasileiro. In: Brazilian Symposium on Computers in Education. Porto Alegre: SBC, 2021. p. 1–10.

STOLL, Stephanie et al. Sign language production using neural machine translation and generative adversarial networks. In: British Machine Vision Conference. Durham: BMVA Press, 2018. p. 1–12.

SUTSKEVER, Ilya; VINYALS, Oriol; LE, Quoc V. Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems. Red Hook: Curran Associates, 2014. p. 3104–3112.

VASWANI, Ashish et al. Attention is all you need. In: Advances in Neural Information Processing Systems. Red Hook: Curran Associates, 2017. p. 5998–6008.