Improving Translation from Portuguese to Brazilian Sign Language with Speech-to-Text Integration through Natural Language Processing

1st Andrew Flores Brongar
Research Group in Computer Science Education
Federal University of Rio Grande
Rio Grande, Brazil
andrewbrongar@gmail.com

2nd Jordan Zitzke Pinho
Research Group in Computer Science Education
Federal University of Rio Grande
Rio Grande, Brazil
jordan.zitzke.rs@gmail.com

3rd Rosa Vicari
Department of Theoretical Informatics
Universidade Federal do Rio Grande do Sul, Brazi
Porto Alegre, Brazil
rosa@inf.ufrgs.br

4thGisele Moraes Simas
Research Group in Computer Science Education
Federal University of Rio Grande
Rio Grande, Brazil
giselesimasrs@gmail.com

5th Regina Barwaldt
Research Group in Computer Science Education
Federal University of Rio Grande
Rio Grande, Brazil
reginabarwaldt@furg.br

***Abstract*—This innovative practice full paper focuses on developing a Portuguese-to-Libras translation model designed to facilitate the participation of deaf students in academic conferences in the absence of sign language interpreters. The pursuit of remote solutions, such as virtual translators for simultaneous interpretation from Portuguese to Libras, is intended to facilitate the inclusion of deaf students in higher education during the pandemic, despite the presence of bureaucratic barriers and interpreter shortages. The education of deaf individuals presents a distinctive set of challenges that necessitate the development of innovative technological and methodological approaches to enhance inclusion and promote academic success. Some studies have investigated these issues and put forward proposals for improving the educational experience of deaf individuals. The overarching objective of the research was to apply natural language processing (NLP) techniques, utilizing the Transformer deep learning neural network architecture, to translate Brazilian Portuguese into Libras, thereby facilitating real-time translation and emulating the actions of human interpreters through a virtual character. The specific objectives were to create a dataset comprising Portuguese sentences and their corresponding Libras representations, develop a speech-to-text algorithm, identify and adapt suitable models for translation, and evaluate the performance of translations and the integration of speech-to-text with the NLP model. These objectives were established to ensure accuracy and efficacy in translation and system integration. This research project entailed the construction of an automated translation model from Portuguese to Libras, based on the T5 architecture from Hugging Face. Additionally, a speech recognition system was developed to transcribe audio into text. The translation accuracy, evaluated using the BLEU metric, was 75.32%, which was compared with the glossed sentences in Libras used as the linguistic corpus. However, the model encountered challenges with diacritics and specific vocabulary. The simultaneous interpreter, which combined speech-to-text with the NLP model, demonstrated promising results in real-time translations to Libras.**

***Index Terms*—communication; students with disabilities; inclusivity; sign language; deep learning.**

The authors thank the Coordination for the Improvement of Higher Education Personnel (CAPES).

I. Introduction

The World Health Organization (WHO) has estimated that by 2024, approximately 5% of the global population will require rehabilitation services due to disabling hearing loss. It is projected that by the year 2050, one in ten individuals will be affected by this condition [1]. The advent of the SARS-CoV-2 pandemic has precipitated a surge in remote university events, thereby engendering the concept of utilizing virtual translators for simultaneous interpretation from Portuguese to Libras, thereby promoting inclusion for deaf students in higher education [2].

In light of the challenges faced by the deaf community, particularly in comprehending technical terminology in academic contexts, as highlighted by Granada et al. [3], it is imperative to adopt innovative methodologies to enhance the inclusion of Libras users. Prior research, including studies by Eryiğit et al. [4] and Veríssimo et al. [5], has delved into the potential of Natural Language Processing (NLP) techniques for the simultaneous interpretation of Sign Language texts.

Automatic speech recognition can be described as a technology that enables a machine to identify and transcribe into text the content spoken by a human [6]. Natural language processing (NLP) applications use language models to estimate the probability of sequences of words, characters, or bytes in natural language, looking for patterns that aid in the syntactic and semantic understanding of the text. NLP has moved closer to machine learning due to the large amount of data generated daily, driving the development of automatic systems and intelligent solutions in text analysis, including recent projects in the Portuguese language [7].

Considering the viability of these approaches to facilitate communication between Libras users and Portuguese speakers, the objective of this work was to apply NLP techniques, based on the Transformer deep learning neural network architecture, to translate from Brazilian Portuguese to Brazilian Sign Language (Libras).

The remainder of this paper is organized as follows: Section II presents a theoretical framework that examines the impact of technological advancements and social inclusion initiatives on the deaf community. Section III outlines the methodology employed in developing our machine translation system. This includes the creation of a bilingual dataset, the training of the NLP translation model, and the integration with a speech-to-text algorithm. Section IV presents the results and discussion. This section offers a performance evaluation of the developed models. It also analyses the implications of these models for educational settings. Finally, it addresses the limitations of the models while proposing future research directions. Section V concludes the paper. It summarizes our key findings.

II. Theoretical Framework

In this section, we present the theoretical basis for this study, which is divided into three main parts. The first part explores the role of technology in promoting social inclusion for the deaf community, highlighting how technological advances have improved communication and access to information. The second part focuses on current technologies and recent advances in assistive devices that enhance the autonomy and independence of deaf individuals. Finally, we discuss quality assessment techniques used in machine translation and speech recognition systems, emphasizing their importance in evaluating the effectiveness of the models developed.

A. Technology and Social Inclusion: Advances for the Deaf Community

Technological advances have been instrumental in facilitating the social inclusion of deaf individuals, offering avenues for communication and access to information [8]. Moreover, the captioning of videos, particularly those directed at the deaf community, has enabled them to engage with audiovisual content more inclusively [9].

Translation applications such as VLibras, Hand Talk, and ProDeaf have helped deaf people communicate in different languages, reducing language barriers when sign language interpreters are not available [10, 11]. In addition, adapted home devices such as sound beacons and audio amplifiers have enabled deaf people to receive important information in their environment.

In conclusion, technology has played a key role in promoting the social inclusion of Deaf people by providing effective means of communication and access to information, and it is essential to continue to invest in and adopt technological solutions to ensure equal opportunities for this community [8].

B. Existing Technologies and Advances

Assistive technology resources range from simple items such as pencils with thicker grips to advanced computerized systems designed to provide independence and autonomy for people with disabilities [12]. In the educational context of the Deaf community, computers and the Internet play an important role, enabling activities such as text production and Internet access through alternative forms of communication, such as Libras-based applications.

VLibras, a partnership between the Ministry of Planning, Development, and Management from Brazil and the Federal University of Paraíba (UFPB), comprises a set of Open Source tools for translating texts, audio, and videos into Libras [13]. However, studies have identified discrepancies between words and signs, indicating the need for improvement in sign regionalization and linguistic structure [13].

Natural Language Processing (NLP) is a Computer Science area that aims to extract more comprehensive representations and meanings from texts written in natural language [14]. This discipline deals with complex challenges such as anaphors and ambiguities and is essential for understanding human language and applied in various language technologies [15]. NLP applications are based on language models that estimate the probability of word, character, or byte sequences in a natural language [16].

Machine Translation (MT) is a relevant subfield of NLP that involves translating text from one natural language to another using software. For this purpose, in-depth linguistic knowledge of the languages involved is utilized, such as translation grammar and multilingual dictionaries [17]. The Transformer deep learning model represents a notable revolution in neural network architectures for natural language processing tasks such as machine translation, text summarization, and text generation. Its attention mechanism allows the model to focus on different parts of the input sequence during processing, promoting a global understanding of context [18].

C. Quality Assessment Techniques

The Bilingual Evaluation Understudy (BLEU) and Word Error Rate (WER) metrics are widely recognized and employed to evaluate the quality of machine translation systems [19]. These metrics enable researchers to quantify and compare the performance of different machine translation methods or recognition systems against correct reference translations or transcriptions, respectively. This approach is crucial for assessing the effectiveness and accuracy of models developed for sign language translation and interpretation [20].

III. METHODOLOGY

This section presents the methodology employed in this study, detailing the processes and techniques utilized to develop our simultaneous interpreter for real-time Portuguese-to-Libras translation. First, we delineate the steps involved in data preparation, including the collection, processing, and formatting of the dataset. Subsequently, the training process of the NLP translation model is discussed, with particular emphasis on the specific algorithms and parameters utilized to enhance performance. The integration of the speech-to-text algorithm with the translation model is then elucidated, demonstrating how these technologies collaborate to facilitate real-time translation. Ultimately, the section concludes with a discussion of the evaluation metrics used to assess the effectiveness of the developed system.

A. Machine Translation and Data Preparation

To build a bilingual corpus, we used data extracted from the AC/DC project [21] and the Wikipedia article database, accessible through Wikidata. After collecting the texts, processing was carried out to remove items that were not significant to the sentence and to reduce the text, limiting it to 42 characters, and avoiding truncation during tokenization [5].

The sentences were saved in a CSV file with a column identifying the language of the sentence. Subsequently, the sentences were translated using the library imported into the Python language called "vlibras_translate"[1], a specific module for translating between Brazilian Portuguese (PT-BR) and Brazilian Sign Language (Libras). After translation, the sentences were saved in the same CSV file, but in a separate column for the translated version.

The preparation of data for the creation of a glossary in Libras, as illustrated in Figure 1, entails a series of fundamental steps. Firstly, it is essential to acquire an extensive set of sentences that will serve as the basis for the glossary. Next, the data undergoes meticulous processing, in which any duplicate or empty lines are removed. Furthermore, a character limitation of 42 is applied, a crucial pre-processing measure to ensure the efficiency of the model and its ability to handle a variety of inputs optimally.

Once the sentences have been processed, any grammatical errors need to be corrected, including diacritics and punctuation. This correction phase contributes to the quality and accuracy of the final glossary. Next, the pre-processed sentences are translated into Libras, maintaining their original grammatical structure, using the vlibras-translate library in Python. This step is crucial to ensure that the translations are faithful to the meaning of the original sentences in Portuguese.

Finally, the translated sentences are stored in a CSV file, prepared for subsequent processing in the translation stage, where appropriate models and tokenizers will be used. This comprehensive process culminates in a bilingual dataset.

To use the dataset, it was necessary to adapt the data into a ".JSON" (JavaScript Object Notation) format with the following structure: "id": "0","translation": "pt": "Isto é couro puro"[2], "lb": "ISTO SER COURO PURO".

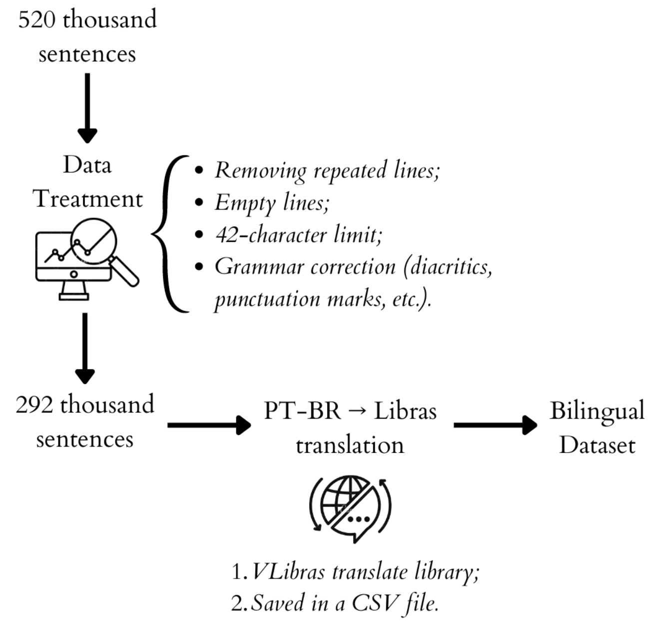


Fig. 1. Data Preparation

B. NLP Translation Model Training

Our sequence-to-sequence (Seq2Seq) NLP translation model was built using an approach guided by the Hugging Face documentation and guidelines [22]. We used the T5 (Text-to-Text Transfer Transformer) architecture. The T5 is a Transformer Deep Learning neural network architecture that provides a unified approach for various tasks, including machine translation. The parameters used are listed below.

- output_dir="./pt_to_Libras_model". Specifies the directory where the trained models and training results will be saved. In this case, the models will be saved in the "pt_to_Libras_model" directory.
- evaluation_strategy="epoch". Sets the evaluation strategy at the end of each epoch. This means that the model performance evaluation will occur once every complete training cycle.
- learning_rate=2e-5. Establishes the learning rate for the optimizer used during training. In this case, the learning rate is set to 2e-5.
- per_device_train_batch_size=16. Defines the batch size per device during training. Each device (e.g., GPU) will process 16 samples simultaneously during each training step.
- per_device_eval_batch_size=16. Similar to the previous parameter, but refers to the batch size during model evaluation.

[1] https://pypi.org/project/vlibras-translate/

[2] This is pure leather

- weight_decay=0.01. Introduces a weight decay term as a form of regularization during training.
- save_total_limit=3. Limits the total number of checkpoints saved during training. In this case, up to 3 checkpoints will be retained.
- num_train_epochs=70. Specifies the total number of training epochs. The model will be trained over 70 epochs.
- predict_with_generate=True. Enables prediction generation during the evaluation step. This is common in automatic translation tasks.
- fp16=True. Activates the use of half-precision (float16), which can speed up training, especially on GPUs compatible with this precision.

At the end of the training, the trained model and the tokenizer were saved for future use in translations from Portuguese into Libras.

C. Function to Correct Translated Words

The function to correct words takes a string and a list of diacritics as input. The function iterates over each diacritic in the list, finding the position of the first occurrence of the diacritic in the sentence. During the iteration, a loop is initiated to continue searching for the diacritic in the sentence until it is no longer found. If a space is found immediately after the diacritic, the function removes that space from the sentence. The process is repeated until there are no more spaces after the diacritic in the sentence. The function prints useful information during execution, such as the position of the diacritic in the sentence and the parts of the sentence before and after the space removal. The result is the adjusted sentence, without unnecessary spaces after the diacritics. This method is used to enhance the readability of the text, especially when diacritics are present.

D. Audio Transcription

The main idea behind the implemented speech recognition system is to capture the user's voice through a microphone and convert it into text, in other words, transcribe it so that the NLP model can utilize it in text format. The speech recognition system was implemented in Python. We utilized the following libraries:

- Threading. Allowed us to use concurrent programming. Two threads were created, one for capturing audio and another for transcribing it.
- Queue. Used to implement a queue to synchronize the two threads. The capture thread puts audio snippets in the queue using the put() method, while the transcription thread retrieves the audio snippets from the queue using the get() method and processes them.
- Time. Employed to enable control over speaking time.
- Sounddevice. Enables us to record audio in real time.
- SpeechRecognition. Essential for the system, it contains various modules for speech recognition. This library serves as a convenient interface, enabling the use of speech recognition services. In this work, we opted to use the Google service.
- Os. Provides functions for working with file paths and directories. In this case, we used it to indicate the path to the Google API credentials.

The implementation of two functions proved necessary: one is responsible for capturing audio and the other for transcribing it.

- capture_audio function: executed by the thread that captures audio snippets from the microphone and places them in the queue (audio_queue).
- transcribe function: executed by the thread that converts audio snippets from the queue into text and adds them to a text list (text_list).

Finally, we opted to use Google Speech-to-Text API, a cloud service that allows developers to convert audio into text. It supports over 100 languages, including Brazilian Portuguese. Additionally, it has an error rate of 5% for most languages and can be easily integrated into the system.

- audio_data: audio snippet to be recognized.
- language: specifies the language of the recognized text, in this case, Brazilian Portuguese ("pt-BR").

E. Speech-to-Text Integration with the NLP Translation Model

To integrate the two developed solutions effectively, it was essential to refine each algorithm by debugging them. After optimizing the solutions, the algorithms were merged. The resulting code implements a comprehensive system for speech recognition and automatic translation in Python, utilizing the Transformers library for natural language processing, threading for concurrent programming, and "speech_recognition" for speech recognition. The imported libraries are shown below.

- transformers: Imports the necessary functionalities for natural language processing, including tokenization and pre-trained models.
- threading: Library to enable concurrent programming using threads.
- queue: Module to implement queues, used for communication between threads.
- speech_recognition as sr: Imports the library for speech recognition and renames it to 'sr' for ease of use.

F. System Quality Assessment

In the end, evaluations were carried out to check the quality of the automatic speech recognition system, using the Word Error Rate (WER) metric (Equation 1) [23]. WER assigns a score in the range of 1 to 0, where a WER metric closer to 0 indicates a null error rate. The following equation defines the WER metric.

$$WER = \frac{I + D + S}{T}, \tag{1}$$

where I represents the number of inserted words; D, the number of deleted words; S, is the number of substituted words; and T, is the total number of words.

The evaluation of the Portuguese to Libras translation model was realized using the Bilingual Evaluation Understudy

(BLEU) metric [24]. BLEU assigns a score in the range of 0 to 100, where a score of 100 indicates a perfect match between two translations.

IV. Results and Discussion

This section presents the findings yielded by our approach. The following section is organized as follows: Subsection A presents the performance of the NLP model in translating Portuguese to Brazilian Sign Language (Libras). Subsection B then examines the efficiency of the speech recognition system. Subsection C goes on to detail the integration of the speech-to-text system with the NLP translation model, offering insights into its practical application in real-time translation scenarios. Finally, Subsection D explores the limitations observed and proposes potential directions for future research to enhance the system's performance and applicability.

A. Natural Language Processing Model

The methodology adopted to develop the machine translation model produced satisfactory results, although not ideal when using the pre-trained 't5-small' model from the Hugging Face platform with the Seq2Seq approach. Initially, the model met the proposal of translating Portuguese grammar into Libras grammar (gloss), but faced limitations when reproducing diacritical characters.

To address this issue, the diacritics were incorporated into the tokenizer manually and the model was retrained for 20 epochs using a dataset developed in this study, comprising 120,000 sentences from each grammar. Following the successful incorporation of the diacritics, however, some undue spaces remained in the generated sentences. To address this irregularity, a post-processing function was implemented, as detailed in Subsection III.C.

Furthermore, some words were not translated correctly. Therefore, we retrained the model with a new bilingual dataset, consisting of 91,995 sentences, to improve the accuracy of the translation. Following the final training, the quality of the translation was evaluated using the BLEU metric, resulting in a hit rate of 75.32% after translating 999 sentences. Table 1 provides a summary of 4 of these translations for comparative evaluation. For the reader's convenience, the presented table has been translated into English. However, it should be noted that the underlying model was developed using the Portuguese language and the grammar of Libras (Brazilian Sign Language).

Finally, Table 2 highlights the performance of each model version developed for translating Brazilian Portuguese into Libras, with our final model (Model 5) achieving the highest BLEU value, 75.3%. Table 3 presents a comparison between our translation model, developed in this study, and VLibras[3]. Our model exhibited a significantly faster translation rate, performing more than six times quicker.

[3]https://www.gov.br/governodigital/pt-br/acessibilidade-e-usuario/vlibras

TABLE I
Glossary of Translations

Origin	Reference Sequence	Sequence Returned by the Model
Against the Man with the Golden Gun	AGAINST MAN WITH GOLD GUN	AGAINST MAN WITH GOLD GUN
although it already brought all the ingredients	ALTHOUGH BRING ALL INGREDIENT ALREADY	ALTHOUGH "BRIMG" ALL INGREDIENT ALREADY
it was expected in Bond films	EXPECT BOND FILM	EXPECT BOND FILM
lazy, a film that existed to	LAZY 1 FILM THAT EXIST TO	1 FILM THAT EXIST TO

TABLE II
Comparison of Trained Models

Test	Pre-Trained Model	No. of Sentences Used	BLEU(%)
1	opus-mt-tc-big-en-pt	30,000	18,8
2	T5-small	30,000	56,4
3	Model 1	120,000	70,3
4	Model 2	60,000	73,8
5	Model 3	120,000	69,9
6	Model 4	22,000	71,0
7	Model 5	91,995	75,3

B. Speech-to-Text

As illustrated in Table 4, the WER metric demonstrates sensitivity to uppercase, lowercase, and punctuation. In the initial sentence, the absence of a period in the hypothesis is identified as an error. In the second sentence, the error is attributed to the lack of a period and the beginning of the sentence in lowercase. Conversely, in the third sentence, the metric yields a score of 0.0, indicating a completely accurate transcription.

To address these issues, the sentences were transformed into lowercase using Python's lower method before calculating the WER, as illustrated in Table 5. This eliminates errors related to punctuation and differences between upper and lower case. With these modifications, it was possible to achieve an average

TABLE III
Comparing Translation Methods

Method	Translation Time(s)	Tested Phrases
Translation Model	209	999
VLibras	1335,2	999

TABLE IV
Example of exceptional cases of the WER metric

Reference	Hypothesis	WER (%)
A vida é feita de escolhas**.**	A vida é feita de escolhas	0.16
O cachorro está correndo**.**	o cachorro está correndo	0.33
Albert Einstein foi um físico teórico	Albert Einstein foi um físico teórico	0.0

WER of 0.11. Having established this approach, we proceeded to collect a sample of sentences for testing, obtained using a Sony Pulse headset in an environment free of external noise. Ten sentences were selected for speech-to-text evaluation, each uttered only once, as shown in Table 5.

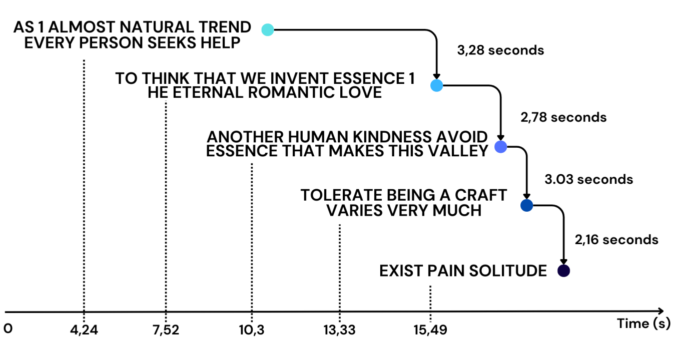
TABLE V
TESTING WER METRICS

Reference	Hypothesis	WER (%)
O sistema está em funcionamento	o sistema está em funcionamento	0.0
Ayrton Senna foi um piloto de Fórmula 1	Ayrton Senna foi um piloto de Fórmula 1	0.0
O píer de Santa Mônica fica em Los Angeles	O que é de Santa Mônica fica em Los Angeles	0.22
Serviços de streaming estão cada vez mais populares	serviços de streamings estão cada vez mais populares	0.12
A seleção brasileira possui cinco copas do mundo	a seleção brasileira possui cinco copas do mundo	0.0
Um carro com mais de trinta anos é isento de IPVA	um carro com mais de 30 anos é isento de IPVA	0.09
A segunda guerra mundial acabou em 1945	a Segunda Guerra Mundial acabou em 1945	0.0
Se exercitar é importante para saúde	se exercitarem importante para saúde	0.33
A gasolina não pode acabar durante a viagem	a gasolina não pode acabar	0.37
Hoje o dia está nublado	Hoje o dia está nublado	0.0

C. Simultaneous Interpreter

Integrating the two technologies developed, Speech-to-Text and the Natural Language Processing Model, it was possible to analyze the quality of the simultaneous interpreter. This interpreter receives the audio obtained from the microphone every 3 seconds, transcribes it, and sends it to the NLP model for translation. Figure 2 shows the sentence translated into Libras grammar.

Fig. 2. Time Diagram - Continuous Sentence Reading



The initial interval, from 0 seconds to 4.24 seconds, represented the time it took to obtain the sentence "AS 1 ALMOST NATURAL TREND EVERY PERSON SEEKS HELP". The next sentence took 3.28 seconds to process. The total time for each sentence can be seen in Table 6, along with the transcription and translation times. To better visualize the results of our approach, an excerpt was read continuously and spoken by a speaker. The captured audio was then transcribed and translated automatically by the system developed. Table 7 shows the results obtained.

TABLE VI
TIME FOR TRANSCRIPTION AND TRANSLATION

Text	Transcription (s)	Translation (s)	Duration (s)
Sentence 1	3,33	0,91	4,24
Sentence 2	2,41	0,87	3,28
Sentence 3	1,90	0,88	2,78
Sentence 4	2,40	0,63	3,03
Sentence 5	1,40	0,76	2,16

TABLE VII
TRANSCRIBED AND TRANSLATED DATA

Original Text	Transcription	Translation by Modelo
Como uma tendência quase natural de todas as pessoas buscarem auxílio.	como uma tendência quase natural de todas as pessoas buscarem auxílio	COMO 1 TENDÊNCIA QUASE NATURAL TODO PESSOA BUSCAR AUXÍLIO
Eu acho que nós inventamos essências, uma delas é o amor eterno, romântico;	acho que nós inventamos essências uma delas e o amor eterno romântico	ACHAR QUE NÓS INVENTAR ESSÊNCIA 1 ELE AMOR ETERNO ROMANTICO
a outra é Deus; a outra é a bondade humana. Inventamos essências que tornam esse vale	a outra bondade humana evitamos essências que tornam este Vale	OUTRO BONDADE HUMANO EVITAR ESSÊNCIA QUE TORNAR ESTE VALE
de lágrimas tolerável. São opiáceos variados em meio a muitas dores.	tolerável são ofícios variados em meio a muitas	TOLERAR SER OFÍCIO VARIAR MEIO MUITO
Existe a dor, a solidão...	existe a dor da Solidão	EXISTIR DOR SOLIDÃO

Khan et al. [25] developed a machine translation model for converting English sentences into equivalent Pakistan Sign Language (PSL). Their model achieved an accuracy reflected by a BLEU score of 78% and a WER metric of 0,10, which are comparable to the results obtained in our study.

The results obtained from this integration have significant implications for the education of deaf students and the broader educational community. By providing real-time translation of spoken language into Libras (Brazilian Sign Language), this system can enhance the accessibility and inclusivity of classroom environments. Deaf students will be able to receive immediate translations of lectures and discussions, allowing them to participate more fully in the educational process.

For teachers, this technology offers a powerful tool to ensure that their instruction is accessible to all students, including those who are deaf or hard of hearing. It can be integrated into live lectures, recorded classes, and educational resources, making it easier for teachers to communicate effectively with all students. This promotes a more inclusive teaching approach and helps teachers to be more responsive to the needs of diverse learners.

Moreover, the ability to transcribe and translate spoken content quickly and accurately supports the development of teaching materials and resources tailored for deaf students. This technology also facilitates better communication between deaf and hearing students, promoting a more collaborative and inclusive learning environment.

Furthermore, implementing this system in educational settings can raise awareness about the needs and capabilities of deaf students, encouraging educators to adopt more inclusive teaching practices. Consequently, this technology has the potential to significantly improve the educational experiences and outcomes for deaf students, fostering greater equity and inclusion in education.

In conclusion, this study illustrates how the integration of speech-to-text and natural language processing technologies can be effectively utilized in educational settings to benefit both teachers and students, particularly in promoting inclusivity and accessibility.

D. Model limitations

The NLP Machine Translation model developed has been shown to have limitations, particularly in the difficulty it encounters when dealing with diacritical characters. This is due to the sensitivity of the tokenizer, even when such characters have been manually included during training. In addition, the insufficient translation of select terms during quality assessment highlights the necessity for ongoing refinements to the model to address specific vocabulary and diverse linguistic structures effectively. This indicates that the model's contextual comprehension could be enhanced.

In considering future limitations, it is essential to take into account the evolving nature of language, the expansion of vocabulary, and potential shifts in grammatical structures over time. This underscores the necessity for updates and the incorporation of more contemporary datasets for training. To address these challenges and enhance the model, we propose investigating advanced natural language processing techniques, utilizing a more diverse and context-specific dataset, and implementing feedback mechanisms for continuous learning.

To adequately address the issue of diacritical characters and specialized vocabulary, it is imperative to implement automated solutions that supersede the necessity for manual adjustments. This can be achieved by developing advanced normalization algorithms that are capable of automatically recognizing and processing diacritical characters, as well as creating dynamic dictionaries that are updated in real-time to include new specific terms. The incorporation of neural networks with attention mechanisms can enhance the model's capacity to navigate diverse linguistic contexts. Continuous training with a spectrum of datasets and automated feedback systems will facilitate the model's ability to learn and adapt continuously. These approaches will guarantee more precise and efficient translation, reducing time and resource expenditure, and enabling the model to adapt to the evolving linguistic landscape in a robust and scalable manner.

V. Conclusion

The methodology employed to develop the machine translation model yielded satisfactory outcomes, although not optimal (BLEU above 80%). Following the identification of issues associated with the handling of diacritical characters, the model underwent a series of refinements, including the introduction of corrections to address the presence of spaces within generated sentences.

A final training session utilizing an alternative dataset resulted in an enhanced hit rate of 75.32%. Nevertheless, challenges persist in the field of machine translation, particularly about linguistic structures and specific vocabulary.

The implementation of the simultaneous interpreter, which combines speech-to-text with a natural language processing model, demonstrated promising results, providing real-time translations for the grammar in Libras.

The developed system provides real-time translations from spoken Portuguese to Libras, thereby bridging a critical communication gap and facilitating enhanced engagement of deaf students in classroom activities, lectures, and discussions. This not only enhances comprehension but also fosters an inclusive learning environment where all students can participate equitably. The planned extension of the study will incorporate direct tests with deaf individuals and students, to conduct a more comprehensive and in-depth evaluation. The collaboration with deaf students will facilitate the acquisition of valuable insights into the practical usefulness of the proposed approaches in inclusive educational settings.

References

[1] World Health Organization. (2024, January 25). Deafness and hearing loss. https://www.who.int/health-topics/hearing-loss

[2] Souza, Q. R., et al. (2020). INCLUSÃO NA UNIVERSIDADE FEDERAL DE RONDÔNIA/BRASIL: análise acerca da percepção de gestores x aluno. Brazilian Journal of Development, 6(5), 25506-25520.

[3] Granada, R. P., Botelho, S., Barwaldt, R., Lussanrriaga, M. D., Gentil, N. Z., & Espíndola, D. (2019). Computational glossary in LIBRAS: an experience in undergraduate program of Information Systems. In 2019 IEEE Frontiers in Education Conference (FIE) (pp. 1-5). doi: 10.1109/FIE43999.2019.9028357

[4] Eryiğit, C., Köse, H., Kelepir, M., & Eryiğit, G. (2016). Building machine-readable knowledge representations for Turkish sign language generation. Knowledge-Based Systems, 108, 179-194. doi: 10.1016/j.knosys.2016.04.008

[5] Veríssimo, V., et al. (2019). A study on the use of sequence-to-sequence neural networks for automatic translation of Brazilian Portuguese to libras. In: Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (pp. 101-108). doi: 10.1145/3323503.3360292

[6] Ferreira, H., et al. (2017). An automatic and dynamic knowledge assessment module for adaptive educational systems. In Proceedings of the 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT) (pp. 517-521). IEEE. doi: 10.1109/ICALT.2017.809

[7] Ferreira, H. H. (2019). Processamento de linguagem natural e classificação de textos em sistemas modulares. (Undergraduate thesis in Computer Science). University of Brasília, Brasília, Brazil.

[8] Nogueira, C. M. I., Carneiro, M. I. N., & da Silva, T. D. S. A. (2018). O uso social das tecnologias de comunicação pelo surdo: limites e possibilidades para o desenvolvimento da linguagem. Revista Pesquisa Qualitativa, 6(12), 470-497. doi: 10.33361/RPQ.2018.v.6.n.12.234

[9] Vieira, P. A., Teixeira, E. N., & Chaves, É. G. (2017). Caminhos do olhar: a movimentação ocular de espectadores surdos durante a exibição de vídeos legendados. Trabalhos em Linguística Aplicada, 56, 493-526. doi: 10.1590/010318138649288277571

[10] Oliveira, R. S., Cruz, R. L. S., & Magalhães, R. J. (2017). Uma análise sobre o aplicativo Vlibras:(im) possibilidade de ser instrumento para efetivação da inclusão social via acessibilidade digital. In: 4º Congresso Internacional de Direito e Contemporaneidade.

[11] Andreis-Witkoski, S. (2020). Problematizando o uso do aplicativo de tradução Hand Talk no ensino da Libras no Ensino Superior. Educação, Ciência e Cultura, 25(3), 81-89. doi: 10.18316/recc.v25i3.6614

[12] Rodrigues, M., & Mill, D. (2019). A utilização do aplicativo hand talk para surdos, como ferramenta de melhora da acessibilidade na educação. Educação e Tecnologias: Experiências, Desafios e Perspectivas, 376-389.

[13] Oliveira, I. F., De Sousa, J. L. T., de Carvalho, L. C., Brito, M. D. O., de Aguiar Freire, K. M., & de Sousa Damasceno, A. C. (2020). A utilização do aplicativo vlibras como forma de ensino e aprendizagem para alunos surdos. Revista Psicologia & Saberes, 9(16), 22-30. doi: 10.3333/rps.v9i16

[14] Indurkhya, N., & Damerau, F. J. (2010). Handbook of natural language processing. Chapman and Hall/CRC. doi: 10.1201/9781420085938

[15] Barbosa, J., Vieira, J. P. A., Santos, R., Junior, G. V. M., MUNIZ, M. D. S., & Moura, R. S. (2017). Introdução ao processamento de linguagem natural usando python. III Escola Regional de Informatica do Piauí, 1, 336-360.

[16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[17] Caseli, H., Freitas, C., & Viola, R. (2022). Processamento de Linguagem Natural. Sociedade Brasileira de Computação.

[18] Ashish, V. (2017). Attention is all you need. Advances in neural information processing systems, 30, I.

[19] Ananthanarayana, T., Srivastava, P., Chintha, A., Santha, A., Landy, B., Panaro, J., ... & Nwogu, I. (2021). Deep learning methods for sign language translation. ACM Transactions on Accessible Computing (TACCESS), 14(4), 1-30. doi: 10.1145/3477498

[20] Naranjo-Zeledón, L., Peral, J., Ferrández, A., & Chacón-Rivas, M. (2019). A systematic mapping of translation-enabling technologies for sign languages. Electronics, 8(9), 1047. doi: 10.3390/electronics8091047

[21] Linguateca. (2024, January 15). Acesso a corpos de português: Projeto AC/DC. https://www.linguateca.pt/ACDC

[22] huggingface. (2022, March 23). huggingface.co/docs/transformers/tasks/translation

[23] Ali, A., & Renals, S. (2018, July). Word error rate estimation for speech recognition: e-WER. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 20-24). doi: 10.18653/v1/P18-2004

[24] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

[25] Khan, Nabeel Sabir; Abid, Adnan; Abid, Kamran. A novel natural language processing (NLP)–based machine translation model for English to Pakistan sign language translation. Cognitive Computation, v. 12, p. 748-765, 2020.