ReCore at SPA-MSLG 2026: Fine-Tuning Transformer Models for Mexican Spanish-to-MSL Gloss Translation

Cristian Misael Torres-Martínez[1,*], Cesar Nathanael Tapia-Cantero[1], Carlos Torres-Orozco[1] and Jorge Antonio Toscano-Lara[1]

[1]*Tecnológico Nacional de México / Instituto Tecnológico de Ciudad Guzmán, Jalisco, México, 49100*

Abstract

Translating written Mexican Spanish into Mexican Sign Language Gloss (MSLG) represents a challenging task in Natural Language Processing due to the structural differences between spoken language and gloss-based sign language representations. In this work, we present the approach developed by the ReCore team for the SPA-MSLG 2026 shared task, focused on bidirectional translation between written Mexican Spanish and MSLG. To address the low-resource nature of the task, we implemented a fine-tuning strategy based on the multilingual No Language Left Behind (NLLB) transformer architecture, adapting the model to the structural characteristics of MSLG representations. Our methodology includes dataset preprocessing, vocabulary adaptation, and transformer-based sequence-to-sequence learning for both Spanish-to-gloss (SPA2MSLG) and gloss-to-Spanish (MSLG2SPA) translation. The proposed approach aims to model syntactic and lexical differences between both languages while improving robustness through subword tokenization strategies. This work contributes to the development of accessible language technologies for the Mexican deaf community and demonstrates the potential of multilingual transformer models for low-resource sign language translation tasks.

Keywords

Translation, Fine-Tuning, NLLB, MSL, MSLG, Transformer

1. Introduction

The communication gap experienced by the deaf community represents a significant social and structural challenge in the Mexican context. According to official statistics, approximately 2.9 million people in Mexico have some form of hearing impairment, of which nearly 1.3 million present total deafness [1]. For this population, Mexican Sign Language (MSL) constitutes their primary means of communication and a central component of their cultural identity. However, legislative reports indicate that only 0.2% of the hearing population in the country possesses the ability to understand or use this language [2]. This asymmetry is further aggravated by the severe shortage of professional interpreters, with only around 42 certified interpreters formally registered nationwide [3].

From a linguistic perspective, MSL is a complete and independent language with its own grammatical structure, rather than a direct representation of spoken or written Mexican Spanish. Its structure relies on a visual-gestural phonology composed of three main dimensions: the segmental matrix (movement and hand usage), the articulatory dimension (handshape, spatial location, direction, and orientation), and non-manual features such as facial expressions, head movements, and body posture [4]. In contrast to Mexican Spanish, MSL presents substantial morphosyntactic differences. Generic articles and plural forms are commonly omitted, prepositions are frequently absent, and verbs do not express tense through suffixes. Additionally, the canonical Subject-Verb-Object (SVO) structure of Spanish is often replaced by a more flexible organization centered around Time, Topic, and Comment [5].

Since MSL is an unwritten language and lacks a standardized written representation, computational language processing research depends heavily on the use of glosses [6, 7, 8]. A gloss is an intermediate

IberLEF 2026, September 2026, León, Spain

*Corresponding author.

cristian.ms.martinez@gmail.com (C. M. Torres-Martínez); nathanael0110@gmail.com (C. N. Tapia-Cantero); carlos.torres9858@gmail.com (C. Torres-Orozco); joantoscano963@gmail.com (J. A. Toscano-Lara)

0009-0009-9980-3383 (C. M. Torres-Martínez); 0009-0002-0340-0790 (C. N. Tapia-Cantero); 0009-0005-4456-9703 (C. Torres-Orozco); 0009-0006-0173-0875 (J. A. Toscano-Lara)

and analytical transcription mechanism that represents signed concepts using written words from the spoken language, typically in uppercase letters, while preserving the morphosyntactic and semantic sequence of the original visual-gestural expression [4]. As a result, translating between written Mexican Spanish and MSL glosses represents a challenging task due to the structural and semantic differences between both language systems.

At a computational level, bidirectional translation between Mexican Spanish and glosses involves transforming natural language into a structured semantic representation. In general terms, the system receives an input sentence and segments it into subword units using tokenization algorithms such as Byte Pair Encoding (BPE), allowing the model to efficiently manage grammatical variations and vocabulary limitations [9]. These tokens are then embedded into a high-dimensional vector space, where the transformer encoder captures global sentence dependencies and extracts relevant semantic information [10]. Finally, the decoder autoregressively predicts the output sequence, reorganizing the syntax according to the structural rules of the target representation. The generated gloss sequence can later be employed for both algorithmic evaluation and future applications involving three-dimensional avatar animation systems [11, 12].

To encourage research efforts in this area, the MSLG-SPA 2026 shared task was introduced as part of IberLEF 2026, an evaluation campaign focused on Natural Language Processing (NLP) for Spanish and Iberian languages [13]. MSLG-SPA 2026 constitutes the first shared task dedicated to bidirectional translation between MSL glosses and Mexican Spanish, providing a standardized evaluation framework for automatic translation systems [14]. The task is divided into two subtasks: Spanish-to-gloss translation (SPA2MSLG) and gloss-to-Spanish translation (MSLG2SPA), aiming to foster the development of computational resources and accessible technologies for Deaf communities in Mexico.

The purpose of this work is to present the methodology developed by the ReCore team for the bidirectional translation task between written Mexican Spanish and MSLG. Through the training and fine-tuning of transformer-based generative models, this work addresses two main subtasks: Spanish-to-gloss translation (SPA2MSLG) and gloss-to-Spanish translation (MSLG2SPA). The proposed approach seeks to contribute to the development of accessible assistive technologies and promote inclusive communication tools for the Mexican deaf community [15].

2. Related work

Research on automatic translation for sign languages has evolved through different computational paradigms. Early approaches were mainly based on Rule-Based Models (RBM), which employed static transfer grammars and explicit dictionaries to reorder words from spoken languages into sign language representations [16, 17, 18]. Although these systems achieved acceptable results in controlled lexical environments, they lacked generalization capabilities and struggled with linguistic variability and polysemy in natural language processing scenarios [19].

The incorporation of deep learning techniques introduced sequence-to-sequence (Seq2Seq) architectures based on Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) units for sign language translation tasks [20, 21]. These approaches improved translation fluency by enabling automatic grammatical learning [22]. However, recurrent architectures suffered from performance degradation when processing long sequences due to the vanishing gradient problem, limiting their ability to capture long-range contextual dependencies. This limitation is particularly relevant in sign language translation, where temporal markers and contextual structures can alter the semantic interpretation of an entire sentence [23].

The introduction of the Transformer architecture significantly improved sequence modeling by replacing recurrence with multi-head self-attention mechanisms capable of processing complete sequences in parallel [10]. Through Query, Key, and Value vector projections, Transformers compute relevance relationships between all sentence elements simultaneously, allowing more effective contextual representation learning. In intermodal translation tasks, self-attention mechanisms provide the flexibility required to address structural asymmetries between spoken language and sign language glosses. This

Table 1
Some Typical Commands

ID	MSLG	SPA
220	PEZ COLOR AZUL MI FAVORITO	Los peces azules son mis favoritos.
749	YO FELIZ MI HIJO LICENCIA-DE-CONDUCIR AL YA TENER	Estoy feliz, mi hijo ya tiene licencia de conducir.
180	SU LIBRO MATEMÁTICAS USTEDES SACAR	Saquen su libro de Matemáticas.
239	YO NECESITAR PAPEL PARA AGUA ABSORBER	Necesito papel para absorber el agua.

enables the encoder to identify and suppress unnecessary grammatical elements, such as articles and prepositions, while allowing the decoder to reorganize sentence structure without relying on manually designed rules [24]. Recent evaluations in the Mexican context have shown that encoder-decoder Transformer architectures consistently outperform previous probabilistic approaches by capturing semantic and morphosyntactic patterns more effectively in gloss-based representations [6].

Despite the effectiveness of Transformers, training these architectures from scratch requires large-scale corpora that are rarely available for sign languages, which are generally considered low-resource environments [25]. To mitigate overfitting and data scarcity limitations, recent research has increasingly relied on transfer learning strategies using large pre-trained multilingual models that already encode generalized linguistic representations [26, 27]. Following this paradigm, our work adopts the NLLB200 model developed by Meta, a multilingual Transformer architecture optimized for translation across more than 200 languages, including low-resource scenarios [28]. To adapt the model to MSLG, the methodology follows the fine-tuning and language adaptation strategies proposed by [29]. In particular, a dedicated token for MSLG was incorporated into the tokenizer vocabulary, and its vector representations were initialized within a latent space previously calibrated using Spanish distributions. This strategy reduces stochastic initialization effects, accelerates convergence during training, and improves bidirectional mapping between structurally asymmetric language representations [29].

Another important challenge in gloss-based translation systems is the management of out-of-vocabulary (OOV) terms caused by the compact nature of gloss notation and the presence of low-frequency lexical forms [30]. Traditional word-level tokenization often increases data sparsity, negatively affecting model generalization when unseen terms appear during inference. To address this issue, Natural Language Processing research commonly employs subword segmentation algorithms such as Byte Pair Encoding (BPE) [9]. BPE iteratively decomposes words into statistically recurrent subword units, enabling the decoder to reconstruct and interpret previously unseen lexical forms from known linguistic patterns [31]. This strategy improves autoregressive text generation stability and contributes to producing cleaner outputs suitable for formal evaluation using standard translation metrics [12].

3. Methodology

For the development of this work, the methodology proposed in the tutorial provided by [29] was used as a reference framework, adapting its multilingual fine-tuning strategies to the bidirectional translation task between MSLG and written Mexican Spanish.

3.1. Dataset description

The training dataset provided by the shared task organizers consists of 491 aligned translation instances. Each record contains a sentence in written Mexican Spanish together with its corresponding MSLG translation, as well as a unique identifier associated with the instance. This bilingual structure enables the development of bidirectional translation systems between Spanish and MSLG representations.

The dataset was specifically designed for the Spanish-to-gloss (SPA2MSLG) and gloss-to-Spanish (MSLG2SPA) translation tasks, providing parallel sentence pairs suitable for supervised Neural Machine Translation (NMT) training. Table 1 presents a few examples of the instances included in the dataset.

Table 2
Training hyperparameters

hyperparameter	Value
Training instances	491
Batch size	16
Warmup steps	1,000
Training steps	20,000
Maximum sequence length	128 tokens
Weight decay	1e-3
Learning rate	1e-4
Optimizer	Adafactor
Clip threshold	1.0

3.2. Dataset preprocesing

The training dataset provided for the shared task underwent a preprocessing stage prior to the model fine-tuning process. First, the original bilingual dataset was separated into two aligned datasets: one containing sentences in written Mexican Spanish and another containing their corresponding MSLG representations. This separation facilitated the bidirectional translation training process by allowing the model to independently process each translation direction while preserving sentence-level alignment between both languages.

After dataset separation, the text instances were tokenized using the tokenizer associated with the distilled NLLB200 model. This process transformed the textual sequences into numerical token representations compatible with the Transformer architecture. Tokenization plays a fundamental role in NMT systems, as it enables the model to efficiently process variable-length sequences while preserving semantic and syntactic information. Additionally, subword segmentation strategies allow the system to better manage low-frequency terms and reduce the impact of OOV words, which are particularly common in gloss-based representations due to their compact and domain-specific notation.

3.3. Vocabulary adaption

The model used in this work is a destilled version of the NLLB200 model, which includes a multilingual tokenizer containing the vocabulary and linguistic representations associated with the 200 languages supported by the architecture. Since MSL glosses are not part of the original tokenizer vocabulary, an adaptation process was required to incorporate this new linguistic representation into the model.

To address this limitation, a dedicated token corresponding to MSL glosses was added to the tokenizer vocabulary. The purpose of this additional token was to provide the model with a specific latent representation capable of storing the linguistic patterns associated with the gloss language during training. Instead of initializing this token randomly, its embedding weights were initialized using the representations of an already existing language within the tokenizer that shares structural and lexical similarities with MSL glosses, namely Spanish. This initialization strategy allows the model to begin training from a semantically informed representation space rather than from purely stochastic values.

Although the newly introduced token initially shares the same embedding distribution as the Spanish token, the fine-tuning process progressively enables the model to distinguish the unique morphosyntactic and semantic characteristics of the MSLG. This adaptation mechanism contributes to improving model convergence and generalization capabilities during translation tasks involving low-resource linguistic representations.

3.4. Fine-Tuning

Fine-tuning refers to the process of adapting a pre-trained model for specific tasks. This process is fundamental in deep learning, particularly in the development and optimization of AI models [32].

Table 3
Training Environment Specifications

Component	Specification
GPU	NVIDIA GeForce RTX 4060 Ti
VRAM	16 GB
CPU	Intel Core i7-12700
CPU Speed	4.9 GHz
CPU Cores	20
RAM	32 GB
Operating System	Windows 10

Under this premise, the objective of fine-tuning the distilled NLLB200 model is to enable it to learn a new language and perform bidirectional translation between Mexican Spanish and MSLG.

For the fine-tuning stage, the preprocessed dataset and the adapted tokenizer were used to train the distilled NLLB200 model for bidirectional translation between written Mexican Spanish and MSL gloss. The training process was designed to enable the model to learn both Spanish-to-gloss (SPA2MSLG) and gloss-to-Spanish (MSLG2SPA) translation tasks within the same multilingual framework.

The hyperparameters employed during training are presented in Table 2. The training batch size was set to 16, and the maximum sequence length was limited to 128 tokens, primarily to avoid overloading the available GPU resources of the hardware used for experimentation. Additionally, the training configuration included 1,000 warm-up steps and a total of 20,000 training steps, which was sufficient to reduce the loss while also preventing overfitting on the given training dataset. Furthermore, the weight decay was set to 1e-3 and the learning rate to 1e-4. These values were intentionally kept low because NLLB200 is a pre-trained model; therefore, assigning excessively high values to the learning rate or weight decay could negatively affect the pre-trained token weights. Lastly, the optimizer selected for training was Adafactor, which determines how gradients are used to update the model weights, while the clip threshold limits the maximum gradient magnitude to prevent anomalous data from excessively altering the weights.

For the training cycle, this work was developed in a virtual environment using Python v3.11, specifically employing the Transformers 4.33 and Datasets libraries from Hugging Face [33], the base model and tokenizer downloaded for the fine-tuning is the following: 'facebook/nllb-200-distilled-600M'. The experiments were conducted on the hardware detailed in Table 3. Overall, the complete fine-tuning process required approximately six hours of training time.

At the end of the training stage, a fine-tuned model capable of translating from written Mexican Spanish to MSLG and vice versa was successfully obtained. The resulting model was named NLLB100, where NLLB refers to the base model employed and 100 represents the 100% of instances from the training dataset used during the fine-tuning process. The model was subsequently used to generate translations for the shared task test datasets, and the produced outputs were submitted to the competition organizers for evaluation.

3.5. Evaluation metrics

The proposed model was evaluated using standard Machine Translation (MT) metrics, selected according to the characteristics of each translation direction.

BLEU (Bilingual Evaluation Understudy) [34] measures the similarity between machine-generated translations and reference translations using modified n-gram precision combined with a brevity penalty:

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

where BP is the brevity penalty, p_n is the modified n-gram precision, and w_n is the weight for each

Table 4
Automatic evaluation results for the MSLG2SPA subtask.

Model	Task Score	BLEU	BLEU z-score	METEOR	METEOR z-score	ChrF	ChrF z-score	COMET
NLLB100	-0.1167	22.2298	-0.2659	0.4428	-0.0953	52.5667	0.0958	0.7433

n-gram order.

METEOR [35] evaluates translations by aligning generated and reference sentences using exact matches, stemming, and synonym matching, incorporating both precision and recall:

$$F_{\text{mean}} = \frac{10PR}{R + 9P}, \tag{2}$$

where P is unigram precision and R is unigram recall.

chrF [36] is a character n-gram F-score particularly useful for capturing morphological variations and partial lexical matches:

$$\text{chrF}_\beta = (1 + \beta^2)\, \frac{\text{chrP} \cdot \text{chrR}}{\beta^2\, \text{chrP} + \text{chrR}}, \tag{3}$$

where chrP and chrR are character n-gram precision and recall respectively, and β=1 weights both equally.

COMET [37] is a neural evaluation metric based on pretrained multilingual language models that estimates translation quality by jointly considering the source sentence, the hypothesis, and the reference:

$$\text{COMET}(x, y, r) = f_\theta(x, y, r), \tag{4}$$

where x is the source, y the hypothesis, r the reference, and f_θ a neural regression model. COMET was applied only to the MSLG2SPA direction, since MSL gloss sequences do not constitute fully natural language sentences and therefore do not support the fluency-oriented semantic evaluation that COMET requires.

For system ranking, all metric scores were standardized using z-score normalization across participating systems, and a global Task Score was computed as the arithmetic mean of the normalized values.

4. Results

This section presents the results of the experimental study. The NLLB100 model was evaluated in both translation directions: MSLG2SPA (gloss-to-Spanish) and SPA2MSLG (Spanish-to-gloss).

4.1. MSLG2SPA — Gloss-to-Spanish

Table 4 presents the results for the MSLG-to-Spanish direction. The negative Task Score indicates that the system performed slightly below the field average. A BLEU score of 22.2 suggests that few word sequences exactly match the references, which can be linked to cases where the model used non-existent words or made grammatical errors. It should be noted that the most frequent errors made by this model are those related to semantic inconsistencies, since the sentences are grammatically valid but lack logical sense. Nevertheless, the COMET score of 0.743 and the METEOR score of 0.4428 indicate that the model captures the general topic and employs synonyms that, although not exact, convey meaning similar to the reference translations.

Both results exhibit hallucinations in the form of invented words and spelling errors. In the MSLG2SPA direction, invented verbs and adjectives closely resemble existing Mexican Spanish words, as in *Mi hermano es reporteador en la TV* or *No me arrepudo de nada*, where the intended forms “reportero” and “arrepiento” can be inferred. In other cases the intent is not recoverable, as in *Yo comprigo ya* or *Las próximas inviernes mi casa será muy poderosa.* This behavior is attributed to the limited training data available.

Table 5
Automatic evaluation results for the SPA2MSLG subtask.

Model	Task Score	BLEU	BLEU z-score	METEOR	METEOR z-score	ChrF	ChrF z-score
NLLB100	0.4218	17.4557	0.7722	0.4253	0.2112	55.2070	0.2819

4.2. SPA2MSLG — Spanish-to-Gloss

Table 5 shows the results for the Spanish-to-MSLG direction. The positive Task Score of 0.4218 indicates that this model performed above the competition average. Note that COMET was not computed for this subtask, as MSLG sequences do not constitute fully natural language sentences and therefore do not support the fluency-oriented semantic evaluation that COMET requires [37].

More than half of the generated translations are well-formed and respect MSL conventions, even though the BLEU score of 17.46 is relatively low. This discrepancy is explained by character-level errors such as *HUVO* instead of *HUEVO*, which lower BLEU but are penalized less by the higher chrF score. Hallucinations also appear in this direction, for example *MOSLIMÁN ORAR MECITÁ*, which seems to indicate that someone is praying in a mosque. The superior performance in SPA2MSLG compared to MSLG2SPA is likely due to the lower structural complexity of gloss sequences relative to Mexican Spanish, which requires agreement in conjugations and gender.

5. Conclusions

In the present work, one MT model is developed, capable of translating from Mexican Spanish to MSLG and vice versa. The proposed model based on Meta's NLLB200, shows contrasting performances on both subtasks translation results, with the translation focused on translating from Mexican Spanish to MSLG achieving better results with the proposed dataset. This may be due to the fact that MSLG has a lower complexity that makes it easier to generate than Mexican Spanish, which requires agreement in conjugations and gender. Both translation results exhibit hallucinations, with the MSLG2SPA model inventing Spanish lexicon and the SPA2MSLG model inventing MSLG signs. Overall, the results are encouraging, considering the small sample size. It is hoped that these advances will inspire progress in the field and strengthen inclusivity and technological accessibility. As future work, the expansion of the training set could be proposed, in order to better identify patterns and thus improve translation quality.

6. Acknowledgments

The authors gratefully acknowledge the financial support provided by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), which contributed to the development of this research.

7. Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to improve grammar, spelling, and academic writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] INEGI, Población con discapacidad o limitación en la actividad cotidiana por entidad federativa y tipo de actividad realiza según sexo, 2020, https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=Discapacidad_Discapacidad_02_2c111b6a-6152-40ce-bd39-6fab2c4908e3&idrt=151opc=t, 2020.

[2] Cámara de Diputados, Aprueban reformas para que personas con discapacidad auditiva reciban educación bilingüe en lengua de señas, 2021. Boletín No. 5854 de la Cámara de Diputados del Congreso de la Unión de México.
[3] E. Godoy, Los intérpretes de lengua de señas mexicana intentan afrontar la discriminación por discapacidad, Equal Times (2015).
[4] M. Cruz-Aldrete, Gramática de la lengua de señas mexicana, Ph.D. thesis, El Colegio de México, 2008.
[5] C. A. M. Flores, M. P. Westgaard, L. E. Dellamary, M. C. Aldrete, M. R. G. Ramírez Barba, Diccionario de Lengua de Señas Mexicana de la Ciudad de México, Instituto de las Personas con Discapacidad de la Ciudad de México (INDEPEDI), 2017.
[6] D. V. Lara-Ortiz, R. Q. Fuentes Aguilar, I. Chairez, Design of a neural transformer for spanish to mexican sign language automatic translation/interpretation, Network: Computation in Neural Systems 36 (2025) 45–68.
[7] C. M. Torres-Martinez, V. A. Huerta-García, J. A. Trejo-Sánchez, J. E. Molinar-Solis, D. Fajardo-Delgado, Sistema de traducción automática del español mexicano a glosa de la lengua de señas mexicana mediante aprendizaje profundo, Revista Ingeniantes 12 (2025) 214–222.
[8] D. V. Lara-Ortiz, R. Q. F. Aguilar, I. Chairez, Design of a neural transformer for spanish to mexican sign language automatic translation/interpretation, Network: Computation in Neural Systems 37 (2026) 206–232. doi:10.1080/0954898X.2024.2435495, pMID: 39663581.
[9] R. Sennrich, B. Haddow, A. Birch, Improving neural machine translation models with monolingual data, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 123–129.
[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, volume 30, 2017, pp. 5998–6008.
[11] J. A. Navarrete López, Reconocimiento automático de señas dinámicas de la Lengua de Señas Mexicana, Master's thesis, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), 2025.
[12] B. Saunders, N. C. Camgöz, R. Bowden, Continuous 3d multi-channel sign language production via progressive transformers and mixture density networks, International Journal of Computer Vision 129 (2021) 2113–2135.
[13] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS. org, 2026.
[14] A. Y. Rodríguez-González, D. Fajardo-Delgado, M. Á. Álvarez-Carmona, M. G. Sánchez-Cervantes, Á. Díaz-Pacheco, R. Aranda, S. Fajardo-Flores, C. H. Martínez-Rodríguez, Overview of mslg-spa at iberlef 2026: Bidirectional translation between mexican sign language glosses and spanish, Procesamiento del Lenguaje Natural 77 (2026).
[15] R. E. Cuevas Valencia, G. A. Salgado Martínez, Aplicación móvil basada en deep learning para la inclusión educativa de personas sordas: Reconocimiento y traducción de la lengua de señas mexicana, RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo 15 (2025).
[16] O. Pichardo-Lagunas, B. C. Martínez-Seis, Resource creation for automatic translation system from texts in spanish into mexican sign language, Research in Computing Science 100 (2015) 129–137.
[17] O. Pichardo-Lagunas, L. Partida-Terrón, B. Martínez-Seis, A. Alvear-Gallegos, R. Serrano-Olea, Sistema de traducción directa de español a lsm con reglas marcadas., Res. Comput. Sci. 115 (2016) 29–41.
[18] O. Pichardo-Lagunas, B. Martínez-Seis, A. Ponce-de León-Chávez, C. Pegueros-Denis, R. Muñoz-Guerrero, Linguistic restrictions in automatic translation from written spanish to mexican sign language, in: G. Sidorov, O. Herrera-Alcántara (Eds.), Advances in Computational Intelligence, Springer International Publishing, Cham, 2017, pp. 92–104.
[19] R. San-Segundo, J. M. Montero, R. Córdoba, V. Sama, F. Fernández, L. D'Haro, V. López-Ludeña,

D. Sánchez, A. García, Design, development and field evaluation of a spanish into sign language translation system, Pattern Analysis and Applications 15 (2012) 203–224.
[20] M. Amin, H. Hefny, M. Ammar, Sign language gloss translation using deep learning models, International Journal of Advanced Computer Science and Applications 12 (2021) 312–321.
[21] H. Choi, K. Cho, Y. Bengio, Fine-grained attention mechanism for neural machine translation, Neurocomputing 284 (2018) 171–176.
[22] J. Hernández-Cruz, Traducción de texto en español a texto LSM usando aprendizaje profundo, Master's thesis, Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo, 2019.
[23] N. C. Camgöz, O. Koller, S. Hadfield, R. Bowden, Sign language transformers: Joint end-to-end sign language recognition and translation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10023–10033.
[24] S. Egea Gómez, E. McGill, H. Saggion, Syntax-aware transformers for neural machine translation: The case of text to sign gloss translation, in: Proceedings of the 14th Workshop on Building and Using Comparable Corpora (BUCC 2021), 2021, pp. 18–27.
[25] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef, C. Vogler, M. Ringel Morris, Sign language recognition, generation, and translation: An interdisciplinary perspective, in: Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, 2019, pp. 16–31.
[26] S. Egea Gómez, L. Chiruzzo, E. McGill, H. Saggion, Linguistically enhanced text to sign gloss machine translation, in: International Conference on Applications of Natural Language to Information Systems, Springer International Publishing, Cham, 2022, pp. 172–183.
[27] A. Moryossef, I. Tsochantaridis, R. Aharoni, S. Ebling, S. Narayanan, Real-time sign language detection using human pose estimation, in: Computer Vision – ECCV 2020 Workshops, Springer International Publishing, Cham, 2020, pp. 237–248.
[28] Y. Koishekenov, A. Berard, V. Nikoulina, Memory-efficient nllb-200: Language-specific expert pruning of a massively multilingual machine translation model, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 3567–3585.
[29] D. Dale, How to fine-tune a nllb-200 model for translating a new language, Medium, 2022. URL: https://cointegrated.medium.com/how-to-fine-tune-a-nllb-200-model-for-translating-a-new-language-a37fc706b865.
[30] B. Zhang, B. Haddow, A. Birch, Prompting large language model for machine translation: A case study, in: International Conference on Machine Learning, PMLR, 2023, pp. 41092–41110.
[31] G. Angelova, E. Avramidis, S. Möller, Using neural machine translation methods for sign language translation, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2022, pp. 273–284.
[32] D. Bergmann, What is fine-tuning?, 2024. URL: https://www.ibm.com/think/topics/fine-tuning, [Retrieved 25 may 2025].
[33] S. M. Jain, Hugging face, in: Introduction to transformers for NLP: With the hugging face library and models to solve problems, Springer, 2022, pp. 51–67.
[34] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation, in: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.
[35] S. Banerjee, A. Lavie, Meteor: An automatic metric for mt evaluation with improved correlation with human judgments, in: Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, 2005, pp. 65–72.
[36] M. Popović, chrf: character n-gram f-score for automatic mt evaluation, in: Proceedings of the tenth workshop on statistical machine translation, 2015, pp. 392–395.
[37] R. Rei, C. Stewart, A. C. Farinha, A. Lavie, COMET: A neural framework for MT evaluation, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 2685–2702. doi:10.18653/v1/2020.emnlp-main.213.