A Hybrid Framework for Bidirectional Translation Between Mexican Sign Language Glosses and Spanish

Jose Luis Guerra Contreras[1,*], Felipe de Jesús Trujillo Romero[1]

[1]Universidad de Guanajuato, México

Abstract

Translating Mexican Sign Language Gloss (MSLG) into grammatically correct Spanish presents significant linguistic challenges due to the stark syntactic differences between the two communication systems. MSLG glosses lack conjugated verbs, prepositions, articles, and copulas, typically structuring themselves under a topic-comment format with a highly flexible syntactic order. In this paper, we present a hybrid translation system developed for the IberLEF 2026 evaluation campaign. Our architecture consists of two independent and highly optimized pipelines: (1) a pure Rule-Based Machine Translation (RBMT) engine for the Spanish to MSLG direction, which implements advanced morphosyntactic heuristics over the spaCy NLP library using the medium `es_core_news_md` model; and (2) a semantic grammatical corrector based on Large Language Models (LLM) utilizing the Google Gemini API for the MSLG to Spanish direction. We describe in detail the engineering of the rule engine, including the preservation of interrogative and exclamatory punctuation, dynamic tagging of proper nouns (*dm-*), prefixing of spelled acronyms (#), and systematic relocation of the main verb before the closing sign. Through empirical evaluation using the official IberLEF corpus of 490 parallel pairs, we demonstrate that this hybrid approach minimizes the catastrophic hallucinations typical of purely neural Seq2Seq translators under low-resource scenarios, while maintaining the original semantic fidelity of deaf users. This paper details the architecture, linguistic resources, prompt engineering, and error analysis of the translations.

Keywords

Mexican Sign Language, Gloss Translation, Large Language Models, Rule-Based Machine Translation, Natural Language Processing, IberLEF 2026

1. Introduction

The development of automatic translation systems for sign languages represents one of the most complex and active frontiers in modern Natural Language Processing (NLP). Unlike spoken languages, sign languages are independent visual-gestural linguistic systems with their own grammar, morphology, and three-dimensional syntax [1]. For the computational treatment of these languages, researchers frequently resort to "glosses", which are written textual representations that consecutively represent the manual signs and gestures made by the signer [2]. In this work, we focus on the Mexican Sign Language Gloss (MSLG) and its bidirectional translation to the Spanish language.

Automatic translation between Spanish and MSLG gloss presents asymmetrical challenges due to grammatical disparity [3]. In the Spanish to MSLG direction, the system must eliminate redundant elements for the sign (articles, prepositions, conjunctions), lemmatize complex verb forms to their infinitive, and reorganize syntactic constituents. Conversely, in the MSLG to Spanish direction, the challenge is to generate grammatical connectors from an extremely fragmented text and resolve semantic ambiguities without incurring textual hallucinations [4].

Within the framework of the MSLG-SPA 2026 shared task [5], organized as part of the IberLEF 2026 evaluation campaign [6] —traditionally dominated by direct sequence-to-sequence (Seq2Seq) neural network approaches or Statistical Machine Translation (SMT) [7, 8]— we propose a hybrid approach that drastically divides the two translation directions:

1. **Spanish to MSLG (Pure Rule-Based):** We implemented a deterministic linguistic rule engine based on the syntactic analysis provided by spaCy [9], significantly improved by using the medium

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
jl.guerracontreras@ugto.mx (J. L. G. Contreras); fdj.trujillo@ugto.mx (F. d. J. T. Romero)

`es_core_news_md` model for precise morphological classification. This pipeline dispenses with LLM calls to guarantee ultra-fast translation adaptable to any regular vocabulary without the risk of hallucinations.

2. **MSLG to Spanish (LLM Post-editing):** We employ the Google Gemini API (Gemini 3.5 Flash) [10] configured via highly specialized system instructions to interpret the idiomatic expressions of the gloss, resolve tacit subjects, and generate natural sentences in Spanish without distorting the original meaning of the signer.

To ensure reproducibility and encourage future research, the source code and the complete implementation of our translation pipelines are publicly available at: https://github.com/jlgc10/MSLG_SPA_2026_Communicador-Innato.

2. Related Work

Sign Language Translation (SLT) has historically been addressed through sign recognition pipelines from video, followed by gloss-to-oral text translation [11]. Early MSLG-to-Spanish gloss translation systems utilized transfer grammars based on hand-written deterministic rules. These systems demonstrate high accuracy in closed domains and controlled sentences, but their inability to generalize to complex syntactic variations or dynamic vocabulary limits their adoption in everyday free-conversation contexts.

With the rise of deep learning, models based on attention mechanisms (such as the Transformer architecture [12]) became the standard for machine translation [13]. However, the effective training of these Seq2Seq networks requires large aligned corpora (tens or hundreds of thousands of parallel pairs) [14]. For low-resource languages, such as Mexican Sign Language, datasets are often limited to a few hundred examples. In these "data starvation" scenarios, neural models trained from scratch suffer from severe overfitting and tend to produce catastrophic hallucinations, inventing fluent Spanish words or sentences that bear absolutely no relation to the original signs [15].

Recently, Large Language Models (LLM) with billions of parameters have demonstrated astonishing capabilities in few-shot translation and grammatical post-editing. Instead of training a neural network with a tiny corpus, the modern paradigm proposes using the semantic capabilities of a pre-trained LLM and guiding it through explicit grammatical rules injected into its context (System Instructions). Our approach combines this technique for the MSLG-Spanish translation with a robust classic rule engine for the Spanish-MSLG direction, achieving an optimal balance between semantic adaptability and offline computation speed.

3. Dataset and Resources

3.1. Corpus Statistics

The corpus used in this project corresponds to the official dataset provided for the IberLEF 2026 MSLG-SPA machine translation task. To ensure robust model development, the dataset was partitioned into training and testing subsets without an explicit validation split. The training corpus consists of 490 parallel sentence pairs aligned between Mexican Sign Language Gloss (MSLG) and standard Mexican Spanish. This training set provided the foundational linguistic variations necessary to iteratively refine the prompt engineering rules for the LLM pipeline and to fine-tune the spaCy deterministic heuristics. Finally, the system's performance was benchmarked against two blind test corpora, each containing 245 sentences for their respective translation directions. The final outputs generated for these test sets were processed and formatted into tab-separated values files, organizing the fields by unique sentence identifiers and wrapping the generated strings in double quotes (`"ID"`, `"SystemOutput"`), thereby complying with the strict submission requirements of the shared task.

3.2. MSLG Gloss Conventions

The computational transcription of LSM into glosses uses specific notations to capture non-manual gestures or physical sign inflections:

- **Prefix 'dm-' (Fingerspelling):** Indicates that a word is spelled letter by letter using the manual alphabet (typically proper nouns or neologisms). Our rule engine must process this prefix and convert the token to the corresponding gloss notation (*dm-EDUARDO*).
- **Plus Sign '+':** Represents compound signs where two manual concepts are integrated to form a single semantic unit (for example, *MAESTRO+MUJER* to refer to a female teacher, or *PAPÁ+MAMÁ* to represent parents).
- **Hash Sign '#' (Acronyms):** Indicates rapid spelling or institutionalized acronyms.
- **Hyphen Sign '-' (Single Compound Signs):** Used to denote words that are represented by multiple words in spoken language but constitute a single unique sign gesture in the sign language (e.g., *FIN-SEMANA* instead of *FIN DE SEMANA*).

3.3. Environment and Technological Resources

Software development was carried out in the Python programming language. For the syntactic rule engine, the Natural Language Processing library spaCy was integrated. To ensure correct morphological classification, the Spanish medium model `es_core_news_md` was downloaded and installed, providing high-precision morphosyntactic tagging compared to the small version (`es_core_news_sm`).

For AI-based translation, the Google AI Studio API was used, consuming the `gemini-3.5-flash-medium` model (optimized for latency and structured reasoning). Due to the limitations of the free API quota (15 requests per minute), a concurrency control mechanism was implemented introducing a mandatory 4.1-second delay between each batch processing in the automation script, along with an exponential retry logic with a maximum of 6 attempts and a 120-second timeout limit per request.

4. Proposed Methodology

Our methodological architecture is composed of two independent systems designed specifically to solve the linguistic challenges of each translation direction.

4.1. Spanish to MSLG Pipeline (Pure Rule Engine)

To generate glosses from Spanish text, we opted for a deterministic approach based on pure rules. This approach guarantees that the system never invents signs that do not exist in the input text. The algorithmic logic processes sentences through the following sequential steps, bringing significant computational and linguistic benefits:

1. **Linguistic Analysis with spaCy:** The input Spanish sentence is analyzed morphologically using the `es_core_news_md` model, which generates tokens tagged with their Part-of-Speech (POS) and additional morphological features. The medium-sized model drastically reduces the misclassification of crucial interrogative pronouns compared to smaller models.
2. **Essential Punctuation Preservation:** Unlike traditional gloss engines that discard all punctuation, our algorithm explicitly detects question marks (`¿`, `?`) and exclamation marks (`¡`, `!`). These signs are computationally preserved to denote mandatory non-manual features (facial expressions) in sign language, which are vital for grammatical intent.
3. **Connector Filtering (Stopwords):** A strict list of Spanish stopwords that lack a gestural equivalent in the gloss (articles, prepositions, and conjunctions like *"el", "la", "de", "con", "y"*) was implemented. Words present in this list are discarded. This reduces visual "noise" and correctly emulates the telegraphic nature of sign language glossing without erasing interrogative pronouns that a blind POS-based deletion would remove.

4. **Verbal Lemmatization:** If a token is identified as a verb (VERB or AUX), its infinitive lemma is extracted and converted to uppercase. Proper nouns (PROPN) are prefixed with the fingerspelling tag `dm-`. Standardizes all actions to a timeless semantic base form, mimicking how signs are performed without spoken morphological conjugations.
5. **Acronym Detection (#):** If a word (longer than one character) was already written entirely in uppercase in the original Spanish sentence, it is prepended with the # character. Automates the tagging of institutional acronyms efficiently without requiring a massive pre-loaded dictionary.
6. **Syntactic Reordering (SVO to SOV):** The engine identifies the main verb, extracts it, and inserts it at the end of the token list. This single operation brings the Spanish Subject-Verb-Object default structure into close alignment with the Subject-Object-Verb topological preference of Mexican Sign Language.
7. **Closing Punctuation Alignment:** If the sentence contains closing question or exclamation marks, the relocation of the verb is dynamically adjusted to be placed **just before** the closing sign. Generates syntactically correct and logical outputs instead of placing the verb erroneously outside the question scope.
8. **Isolated Copula Pruning:** The copula *"SER"* is eliminated if it appears as a single isolated word. Prevents the unnatural inclusion of the "to be" verb in LSM glosses, where such copulas are heavily implied by context.
9. **Spacing Cleanup:** String replacements compact the output to remove double spaces.

4.2. MSLG to Spanish Pipeline (Semantic LLM Post-editor)

For the reverse translation (Gloss to Spanish), the problem is fundamentally one of generative semantic generation. Glosses are fragmented and syntactically "dry" texts. A pure rule engine would generate a robotic Spanish without grammatical cohesion. Therefore, the task is delegated to an LLM trained with very strict system instructions that define the translation behavior.

The language model is configured by injecting into its context 9 linguistic translation guidelines oriented to LSM and Mexican Spanish:

- **Rule 1 (No Hallucinations):** The model must strictly translate the semantic meaning of the provided glosses and must not invent extra information, events, or objects that are not present in the input.
- **Rule 2 (Correct Conjugation):** The LLM must infer the correct verbal tense and conjugate the lemmatized gloss verbs according to the tacit or explicit subject of the sentence.
- **Rule 3 (Preposition Handling):** Since glosses lack prepositions (like "de", "con", "para"), the model must logically infer the relationship between nouns and insert the appropriate prepositions to construct a grammatically correct Spanish sentence.
- **Rule 4 (Gender and Number Agreement):** The system must guarantee morphological agreement across all generated articles, nouns, and adjectives in Spanish, accounting for plurals that might only be implied in the gloss context.
- **Rule 5 (Proper Noun Restoration):** Tokens starting with `dm-` must be recognized as spelled proper nouns. The prefix must be removed, and the word must be capitalized appropriately in the final Spanish output.
- **Rule 6 (Idiomatic Adaptability):** If a direct, literal translation sounds unnatural or robotic, the LLM must adapt the translation to reflect natural, colloquial Mexican Spanish while retaining the exact core meaning.
- **Rule 7 (Handling the '+' Symbol):** Instructs the model not to translate the components of the plus sign literally. It must look for idiomatic equivalents: e.g., *MAESTRO+MUJER* becomes *maestra*, *PAPÁ+MAMÁ* becomes *padres*.
- **Rule 8 (Topic-Comment Syntactic Order):** Teaches the LLM to interpret the inversion of the subject and the object in MSLG. The first listed entity is usually the central topic and not necessarily the agent subject of the verb, preventing the LLM from confusing who performs the action.

- **Rule 9 (Object-Verb Grouping):** Resolves the problem where abstract state nouns preceding transfer verbs (like *CALMA PEDIR*) are misinterpreted as subject adjectives. The LLM is forced to treat them as the direct object of the verb (e.g., *"pidió calma"*).

To guarantee robust automated processing free of conversational explanations (Chain of Thought), the LLM was forced to return its definitive translation wrapped in structured XML tags. The Python script parses this result using regular expressions to write it directly to the final file.

5. Experiments and Results

We conducted exhaustive evaluations of our hybrid pipeline on the IberLEF 2026 test set. The Tables below present qualitative analyses of translations generated by our system.

Table 1
Qualitative Results of the MSLG → Spanish Pipeline (LLM)

Source MSLG Gloss	Literal / SMT Translation	Our System Output (LLM)
MAESTRO+MUJER CALMA ELLA YA PEDIR	La maestro mujer calma ella pedir.	La maestra ya nos pidió calma.
dm-SOFÍA SU PAPÁ+MAMÁ APROVECHARSE-DE APOYO	Sofía sus padres aprovecharse del apoyo.	Sofía se está aprovechando del apoyo de sus papás.
¡FELIZ+APLAUSO dm-EDUARDO EXCELENTE CALIFICACIÓN TÚ TENER!	¡Feliz aplauso Eduardo excelente calificación tú tener!	¡Muchas felicidades, Eduardo! Tienes una excelente calificación.

Table 2
Qualitative Results of the Spanish → MSLG Pipeline (Rule Engine)

Source Spanish Sentence	Generated MSLG Gloss (RuleEngine)
¿Qué haces?	¿QUÉ HACER?
¿Entendiste todo?	¿TODO ENTENDER?
Muy bien, ¿qué onda?	MUY BIEN ¿QUÉ ONDA?
¡Qué gran sorpresa, Eduardo!	¡GRAN SORPRESA dm-EDUARDO!
Él mira la TV.	ÉL #TV MIRAR

During the experimental phase, we found that attempting to pre-process the gloss with the rule engine before sending it to the LLM yielded worse linguistic results. The rigid rule engine destroyed context clues. Feeding the LLM directly with the raw preprocessed gloss (cleaning only characters like hyphens and dm-) increased the perceived quality of the output Spanish drastically.

Using the medium model `es_core_news_md` reduced keyword omission errors in short questions to zero, correctly tagging words like "Qué" as pronouns (‘PRON‘) instead of determiners, generating the correct structure ’¿QUÉ HACER?’.

6. Discussion and Error Analysis

Despite the excellent performance of the proposed hybrid architecture, error analysis in the deterministic pipeline reveals the intrinsic limitations of linguistic rule-based systems compared to generative neural ones:

1. **Dependency on Part-of-Speech Tagging:** The Spanish to MSLG rule engine is extremely sensitive to spaCy's classification decisions. In extremely short sentences without context, the model sometimes misclassifies words. Any POS error in the first stage propagates inevitably.
2. **Absence of Semantic Disambiguation in Rules:** Not using an LLM for the Spanish to MSLG direction, the system is unable to disambiguate polysemy. For example, the word *"saco"* (verb to take out) and *"saco"* (noun coat) receive exactly the same syntactic treatment.
3. **API Operational Stability:** For the MSLG to Spanish pipeline, depending on cloud APIs exposes the system to variable latencies and connection failures. The implementation of exponential backoff logic was fundamental.

7. Conclusion and Future Work

We presented a high-performance machine translation architecture for Mexican Sign Language Gloss (MSLG) and Spanish within the framework of IberLEF 2026. For low-resource linguistic tasks where parallel corpora are extremely small (490 sentences), our hybrid approach capitalizes on the best of both worlds: a robust classical deterministic rule engine (RBMT) powered by spaCy `es_core_news_md` to generate impeccable syntactic glosses from Spanish, and a state-of-the-art LLM-based grammatical corrector (Gemini) guided by linguistic prompt engineering to transform disjointed glosses into natural Spanish.

Future work aims to explore the knowledge distillation of large language models into significantly smaller local and open-source models. Through Parameter-Efficient Fine-Tuning (PEFT/LoRA) techniques on a synthetically augmented dataset, we seek to create a completely local, low-latency bidirectional translator. Furthermore, a primary intention for future development is the creation of a comprehensive visual processing project. This proposed system aims to integrate computer vision algorithms and virtual 3D avatars to physically transform MSLG directly from human video feeds into spoken Spanish, and conversely, translate spoken Spanish into dynamically rendered MSLG visual signs, bridging the ultimate gap in real-time accessible communication for the deaf community.

8. Declaration on Generative AI

In the development of the computational system described herein, we acknowledge the use of Generative AI technologies. Specifically, the Google Gemini model was integrated directly into the MSLG→Spanish translation pipeline as the core semantic reasoning and text generation engine. Furthermore, AI-assisted coding tools were utilized during the software development process to analyze spaCy NLP pipelines. All generated content was thoroughly reviewed, and validated to ensure scientific accuracy and compliance with the IberLEF 2026 shared task guidelines.

References

[1] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef, et al., Sign language recognition, generation, and translation: An interdisciplinary perspective, The 21st International ACM SIGACCESS Conference on Computers and Accessibility (2019) 16–31.

[2] A. Müller, S. Ebling, et al., Considerations for meaningful sign language machine translation based on glosses, Frontiers in Artificial Intelligence 5 (2022) 1016283.

[3] K. Yin, A. Moryossef, J. Hochgesang, Y. Goldberg, M. Alikhani, Including signed languages in natural language processing, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 7347–7360.

[4] S. Stoll, N. C. Camgoz, S. Hadfield, R. Bowden, Text2sign: towards sign language production using neural machine translation and generative adversarial networks, International Journal of Computer Vision 128 (2020) 891–908.

[5] 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).

[6] 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.

[7] A. Othman, M. Jemni, Statistical sign language machine translation: from english written text to american sign language gloss, in: International Journal of Computer Science and Network Security, volume 11, 2011, pp. 65–73.

[8] R. San-Segundo, R. Barra, R. Córdoba, L. D'Haro, F. Fernández, J. Ferreiros, J. Macías-Guarasa, J. Montero, J. Pardo, Speech to sign language translation system for spanish, Speech Communication 50 (2008) 1009–1020.

[9] M. Honnibal, I. Montani, spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing, 2017.

[10] G. Team, R. Anil, S. Borgeaud, Y. Wu, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, et al., Gemini: a family of highly capable multimodal models, arXiv preprint arXiv:2312.11805 (2023).

[11] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, R. Bowden, Neural sign language translation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7784–7793.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, volume 30, 2017.

[13] N. C. Camgoz, 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.

[14] M. De Coster, M. Van Herreweghe, J. Dambre, Machine translation from spoken language to sign language using pre-trained language models as encoders, Machine Translation 37 (2023) 1–25.

[15] A. Moryossef, A. Müller, S. Ebling, Y. Goldberg, Data augmentation for sign language gloss translation, in: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages, 2021, pp. 22–32.