A Stochastic Parrot in Searle's Chinese Room: Results from the MSLG–SPA Shared Task at IberLEF 2026

Jordi Porta-Zamorano

Laboratorio de Lingüística Informática, Universidad Autónoma de Madrid, Madrid, Spain

Abstract

We describe UAM_ChineseRoom, a prompt-based system for bidirectional translation between Mexican Sign Language glosses (MSLG) and written Spanish (SPA), submitted to the MSLG–SPA shared task at IberLEF 2026. The system uses an instruction-following large language model whose behavior is constrained by manually refined prompts encoding linguistic heuristics for constituent order, temporal framing, argument structure, gloss normalization, lexical transfer, and Spanish morphosyntactic reconstruction. The two translation directions require asymmetric forms of control: Spanish-to-MSLG translation must suppress Spanish morphology, recover implicit arguments, and reorganize input into compact gloss sequences, whereas MSLG-to-Spanish translation relies more strongly on the model's generative capacity to reconstruct fluent Spanish from compressed symbolic input. In the official evaluation, the system ranked sixth for Spanish-to-MSLG with 21.35 BLEU and fourth for MSLG-to-Spanish with 50.52 BLEU, reflecting both the usefulness of prompting in a low-resource setting and the greater difficulty of producing target gloss order and explicitness. Qualitative analysis shows that the system often preserves temporal framing, argument relations, and regional Spanish choices, but remains vulnerable to errors in gloss selection, event-structure interpretation, pragmatic lexical choice, and nested dependency recovery. We interpret these findings through the image of a stochastic parrot in Searle's Chinese Room: the system performs effective symbolic mediation by combining learned distributional regularities with explicit procedural prompts, yet it remains limited by the use of glosses as textual surrogates for embodied, multimodal signed communication.

Keywords

sign language translation, text-to-gloss, gloss-to-text, Mexican Sign Language, Mexican Spanish, large language models, prompt engineering, stochastic parrots, Searle's Chinese Room

1. Introduction

Machine Translation (MT) for sign languages remains a particularly challenging frontier within Natural Language Processing (NLP). This is primarily due to its low-resource setting, characterized by a scarcity of parallel corpora and standardized linguistic resources. A central difficulty in translating between spoken and signed languages is that they differ fundamentally not only in lexicon, but also in modality and linguistic organization. While spoken languages are primarily linear and auditory-vocal, sign languages are visual-gestural, using the hands, face, gaze, body, and surrounding space simultaneously. Consequently, signed utterances encode grammatical and discourse information through spatial reference, classifier constructions, role shifts, and non-manual markers, such as facial expressions and body posture, rather than relying solely on function words or inflectional morphology. These structural differences directly affect word order, argument structure, aspectual marking, reference tracking, and the expression of topic and focus.

Research from the Sign Language Processing project [1] reframes sign language translation not as a single rigid pipeline, but as a dynamic graph of interconnected conversions across data representations. This framework distinguishes five main representations: raw video, which is information-rich but computationally heavy and privacy-sensitive; pose or skeleton keypoints, which provide a lower-dimensional and more anonymized representation of movement; formal notation systems such as SignWriting [2] or HamNoSys [3]; glosses, which offer sign-by-sign labels at the cost of losing expressive nuance; and standard spoken-language text. These representations are mapped onto a fully connected

IberLEF 2026, September 2026, León, Spain

jordi.porta@uam.es (J. Porta-Zamorano)

https://www.lllf.uam.es (J. Porta-Zamorano)

0000-0001-5620-4916 (J. Porta-Zamorano)

pentagon, in which each node represents a data state or representation and each of the 20 directed edges defines a distinct translation or processing task. Navigating this graph therefore requires an interdisciplinary approach: some edges correspond mainly to computer vision problems, others fall within natural language processing, and the most complex paths require a close integration of both areas.

The present work focuses on two edges of the graph of tasks, text-to-gloss and gloss-to-text translation, within the MSLG–SPA shared task [4] at IberLEF 2026 [5], which addresses the challenge of bidirectional translation between Mexican Sign Language glosses (MSLG) and written Spanish (SPA). While glosses provide a compact symbolic interface between signed and spoken-language data, relying on them introduces clear methodological limitations. Glosses are highly valuable for annotation and computational processing, yet they are not equivalent to the signed language itself. They reduce a simultaneous, multi-channeled visual-gestural signal to a linear sequence of written labels. Consequently, this representation often omits, simplifies, or underspecifies crucial information carried by non-manual markers, spatial indexing, movement modulation, classifier constructions, and the discourse-level signing space.

The present work explores such an instruction-guided approach for bidirectional MSLG–SPA translation. Recent advances in Large Language Models (LLMs) have fueled a growing interest in prompting-based approaches for Machine Translation (MT), particularly in low-resource scenarios where task-specific fine-tuning data is scarce. Rather than relying exclusively on supervised parameter adaptation, prompting guides models through natural-language instructions, in-context demonstrations, heuristics, and explicit linguistic constraints. This paradigm is especially attractive for sign language translation tasks, where scarce domain data can be supplemented by incorporating linguistic expertise directly into the input context. Our prompts encode explicit procedural information concerning constituent order, argument structure, possessive constructions, temporal framing, and lexical composition observed in the training data. By doing so, the system combines data-driven language modeling with symbolic linguistic guidance to actively compensate for the scarcity and structural heterogeneity of the available resources.

From a conceptual perspective, this work presents a hybrid architecture that can be interpreted as a synthesis of Searle's Chinese Room and the stochastic parrot metaphor. The Chinese Room thought experiment, originally proposed by Searle [6], argues that the formal manipulation of symbols according to explicit rules is insufficient for genuine understanding, even when the resulting output mimics that of a competent language user. Conversely, the stochastic parrot metaphor, introduced by Bender et al. [7], offers a non-anthropomorphic view of large language models: fluent linguistic output may emerge from statistical patterning over language data without entailing grounded understanding or communicative intention.

Consequently, the main contribution of this work is twofold. First, we present a prompting strategy for bidirectional MSLG–SPA translation that incorporates linguistically motivated heuristics specifically designed for gloss-based sign language representations. Second, we analyze the implications of prompting-based MT systems through the joint lenses of the stochastic parrot and the Chinese Room, using sign language translation as a case study to examine the boundaries between symbolic manipulation, linguistic competence, and semantic grounding in contemporary LLM architectures.

2. Spanish and Mexican Sign Language

From a typological perspective, Spanish is a predominantly inflectional and generally head-initial Romance language characterized by rich verbal morphology, grammatical agreement, clitic systems, and relatively flexible constituent order. Although the unmarked order of declarative clauses is SVO (Subject-Object-Verb), word order is strongly conditioned by information structure, including topic and focus. Grammatical relations are expressed through a combination of inflectional morphology, function words, cliticization, and syntactic structure, while tense, aspect, mood, evidential or epistemic nuances, and discourse relations may be conveyed through verbal morphology, periphrastic constructions, lexical

choices, and pragmatic context. Spanish is also a pluricentric language, with several national standards and regional varieties distributed across the Spanish-speaking world. This pluricentricity is relevant for translation, since lexical choice, address forms, pragmatic conventions, and some grammatical preferences may vary across regions. In this work, Spanish is considered in relation to Mexican Sign Language (MSL), and the Spanish side of the task is therefore interpreted with particular attention to Mexican Spanish. In addition, information-structure phenomena such as topicalization, focus, regionally variable clitic doubling, and null subjects interact closely with word order and pragmatic interpretation. As a result, translation into and from Spanish requires not only lexical transfer, but also sensitivity to morphosyntactic, discourse-pragmatic, and diatopic variation, often involving substantial restructuring at the syntactic, morphological, and discourse levels.

Mexican Sign Language is the main focus of this work, but it should be understood within Mexico's broader sign linguistic ecology. A major source of linguistic information for MSL is Cruz-Aldrete [8]. MSL belongs historically to the French Sign Language (LSF) lineage and it is linked to American Sign Language (ASL), although it has evolved as an independent sign language with its own grammatical and sociolinguistic characteristics. It is also important to note that MSL is not the only sign language used in Mexico, where Indigenous, local, and community-based sign languages also exist, as the Yucatec Maya Sign Language (YMSL). From a typological perspective, MSL is a visual–gestural language in which meaning is distributed across simultaneous manual and non-manual channels, including handshape, movement, orientation, facial expression, gaze, body posture, and spatial organization [8, 9]. Like many sign languages, it makes extensive use of classifier constructions, spatial indexing, role shift, iconicity, and discourse-driven constituent order, while grammatical distinctions such as aspect, negation, interrogation, emphasis, and referential tracking may be expressed through non-manual markers and movement patterns rather than through overt morphology alone. Information structure, especially topic–comment organization, plays a central role in constituent ordering and interpretation. In this work, MSL is represented through gloss annotations, which provide a useful but partial approximation of signing because they linearize and compress a multimodal linguistic signal into discrete textual labels.

3. Related Work

Previous work on sign-language translation has explored a wide range of approaches, including rule-based transfer [10], statistical machine translation [11], and, more recently, neural machine translation models [12, 13]. A broader survey of machine translation approaches for sign languages is provided by Núñez-Marcos et al. [14].

MTOB (Machine Translation from One Book), introduced by Tanzer et al. [15], is a benchmark designed to evaluate whether large language models can learn to translate a genuinely unseen low-resource language using only human-readable linguistic documentation rather than large parallel corpora. The benchmark focuses on Kalamang, an endangered language with fewer than 200 speakers and virtually no web presence, making contamination from pretraining data unlikely. The task differs from traditional machine translation settings because models must acquire linguistic competence from grammar explanations, bilingual lexicons, and limited examples, resembling second-language learning more than standard data-driven MT. The authors evaluate several prompting and in-context learning strategies with contemporary LLMs and compare them against a human learner who studied the same materials. Although larger models and richer context improve performance, the models still remain below human performance, highlighting the difficulty of true language acquisition from sparse documentation. The work is particularly relevant for low-resource and endangered-language MT because it explores alternatives to large-scale parallel corpora and emphasizes the value of curated linguistic resources produced by field linguists. More broadly, the paper positions itself within recent research on evaluating genuine generalization capabilities of LLMs beyond memorization or web-scale exposure. Unlike many benchmarks that may overlap with pretraining data, MTOB provides a controlled setting for studying adaptation to unseen linguistic systems and therefore contributes to

ongoing discussions about data efficiency, in-context learning, and language acquisition in large models.

4. MSLG–SPA Parallel Corpus

The dataset provided for MSLG–SPA 2026 is designed to support controlled, reproducible, and comparable research on bidirectional translation between Mexican Sign Language Glosses (MSLG) and Spanish (SPA). The corpus consists of aligned sentence-level pairs of Mexican Sign Language gloss sequences and Spanish text sentences. The training split contains approximately 490 aligned sentence pairs. Linguistically, the dataset is composed primarily of short and medium-length everyday sentences rather than extended discourse, including declaratives, interrogatives, imperatives, exclamatives, and coordinated structures. The Spanish side covers frequent communicative domains such as family, education, work, transportation, geography, food, emotions, religion, health, and daily activities.

The corpus exhibits a wide range of morphosyntactic and discourse-related phenomena relevant to machine translation. On the Spanish side, examples include verbal periphrases, possessive constructions, existential clauses, clitic doubling, null subjects, adjectival modification, movement verbs, subordinate structures, support-verb constructions, modality, and temporal-aspectual distinctions. The MSLG side reflects characteristic gloss-level properties such as explicit temporal framing, topic-prominent organization, reduced use of function words, overt pronoun realization, and frequent sentence-initial temporal markers. Perfective or completive interpretations are often marked through the gloss YA, while compounds and lexicalized units are represented using hyphenated glosses.

The dataset also contains numerous examples of structural asymmetry between Spanish and gloss representations. Spanish grammatical information encoded morphologically through agreement, tense, aspect, clitics, or determiners is frequently absent, redistributed, or only implicitly represented in gloss form. Conversely, gloss sequences often privilege information structure and semantic prominence over canonical spoken-language syntax. As a result, the task involves substantial syntactic and discourse-level restructuring rather than straightforward lexical substitution.

In addition, the corpus incorporates several normalization conventions commonly used in sign language gloss annotation. Glosses are represented in uppercase, fingerspelled anthroponyms are marked with the prefix dm-, fingerspelled acronyms are marked with #, and many multiword units appear as compound glosses. Although glosses provide a useful intermediate representation for computational processing, they necessarily linearize and compress a multimodal linguistic signal. Consequently, the dataset constitutes not only a machine translation resource, but also a representation of the broader challenge of mapping between spoken-language text and partially specified symbolic approximations of signed communication.

5. A Prompt-based Translation Architecture

LLM-based work has introduced prompting as an alternative to conventional supervised adaptation, especially in low-resource scenarios where fine-tuning data are scarce. Compared with fully data-driven systems, prompting offers a more interpretable way of injecting linguistic knowledge through explicit instructions and transfer heuristics.

The proposed system can be understood as a prompt-based translation architecture composed of two tightly coupled components: a pretrained large language model and an externalized layer of symbolic procedural guidance encoded in the prompts. In this framework, the prompts are not merely auxiliary instructions, but part of the effective computational mechanism of the system itself. They specify formatting conventions, constituent-order preferences, temporal and aspectual heuristics, lexical transfer strategies, and other linguistically motivated constraints that guide the translation process. The resulting architecture therefore distributes the translation procedure across both the latent multilingual competence of the LLM and the explicit symbolic knowledge embedded in the prompt context.

The initial prompts were generated with GPT-5.4 from the available data by asking the model to propose an effective translation strategy between Spanish and MSLG. They were then manually revised,

refined, and extended through iterative evaluation on the training data. Recurrent errors observed in preliminary outputs were used to adjust the instructions, add linguistic heuristics, and make the translation procedure more explicit. Particular attention was given to phenomena that appeared frequently in the corpus, including tense and aspect marking, implicit and explicit subjects, verb chains, valence and argument structure, and the placement and interpretation of temporal adverbs.

Notably, the final prompt developed for the Spanish-to-MSLG direction is substantially larger and more detailed than the MSLG-to-Spanish prompt. This asymmetry reflects the greater degree of procedural control required to transform morphologically rich Spanish input into constrained gloss representations while preserving topic structure, temporal framing, and conventionalized MSLG ordering patterns. By contrast, the MSLG-to-Spanish direction relies more heavily on the generative and reconstructive capabilities of the language model to infer implicit grammatical information and produce fluent Spanish output. From this perspective, the behavior of the system emerges from the interaction between probabilistic language modeling and procedural linguistic control, occupying an intermediate position between purely neural generation and explicitly rule-guided symbolic processing.

5.1. Spanish to Mexican Sign Language Glosses Prompt

The prompt is designed as a linguistically informed procedural framework for Spanish-to-MSLG translation using an instruction-following LLM. Rather than relying exclusively on statistical generation, it encodes explicit transfer heuristics intended to approximate recurrent structural patterns observed in Mexican Sign Language glosses. The prompt defines the task as the production of normalized MSLG gloss sequences from Spanish input and establishes a controlled output format based on uppercase glosses, linearized gloss ordering, explicit treatment of proper names and acronyms, and suppression of explanatory text. A central design principle is the reinterpretation of Spanish syntax according to discourse-driven and topic-oriented structures characteristic of gloss representations. The prompt therefore introduces rules governing constituent order, especially in verbal chains, movement verbs, topicalized arguments, and infinitival constructions, while also specifying strategies for temporal framing, aspectual interpretation, possession, adjectival placement, and clitic-related phenomena.

A major component of the prompt concerns the treatment of implicit grammatical information in Spanish. Since Spanish frequently omits overt subjects and encodes tense, aspect, modality, and agreement morphologically, the prompt provides heuristics for making relevant semantic information explicit in MSLG glosses. This includes instructions for recovering implicit subjects, handling modal and control verbs, interpreting perfective readings, and deciding when temporal particles such as YA should or should not be produced. Several rules are motivated by information structure: complements may be topicalized, verbs reordered according to argument realization, and topic–comment organization prioritized over strict preservation of Spanish constituent order.

The prompt also incorporates lexical-semantic transfer strategies. It includes guidance for handling genitive constructions, support-verb constructions, movement verbs, clitic doubling, and causative or subordinated structures. Where appropriate, this includes Spanish N_1 *de* N_2 patterns such as *la puerta de la casa* ('the door of the house'), *el color del carro* ('the color of the car'), or *el nombre de mi hermana* ('my sister's name'), which may require reinterpretation as possession, attribute assignment, or a topicalized nominal relation rather than a direct word-for-word rendering. Particular attention is paid to the interaction between syntax and discourse interpretation, especially in cases where Spanish grammatical distinctions do not map directly onto gloss sequences. In addition, the prompt attempts to normalize output through consistent gloss conventions, including capitalization, compounding, hyphenation, and special markers for names and abbreviations.

The prompt also specifies how gender marking should be handled in gloss output. For many feminine common nouns, it uses the marker +MUJER attached to the corresponding masculine gloss, as in *abuela* ('grandmother') → ABUELO+MUJER, *amiga* ('female friend') → AMIGO+MUJER, or *maestra* ('female teacher') → MAESTRO+MUJER. At the same time, the prompt notes lexical exceptions that should remain unchanged, such as *mujer* ('woman'), *mamá* ('mother'), or specialized forms like *enfermera* ('nurse'). Relatedly, proper names of persons are marked with the prefix dm- when appropriate, indicating that

the name should be fingerspelled in the gloss output. The prompt also recognizes that some names, especially those corresponding to well-known historical figures or other widely recognized persons, may have conventionalized name signs of their own rather than requiring fingerspelling. This distinction helps separate ordinary anthroponyms from toponyms, institutions, and other named entities, while also allowing for established signed forms when they are part of the language tradition.

The prompt also makes its preferred sentence template explicit. Its default global order is defined as:

[TIME] [LOCATION/FRAME] [TOPIC] [PARTICIPANT(S)] [PREDICATE] [ASPECT/MODAL/ NEG/EVALUATION].

This template encodes the priority given to temporal framing, discourse organization, and topic–comment structure over direct preservation of Spanish SVO order. The prompt further specifies a fronting hierarchy in which temporal expressions, locations or frames, and topicalized constituents are placed before core participants and predicates whenever they are present.

Conceptually, the overall design combines symbolic linguistic instruction with probabilistic generation. The prompt acts as an externalized rule system that constrains and guides the latent multilingual competence of the language model. In this sense, the architecture resembles a procedural translation manual layered over a generative neural model: explicit heuristics provide structured symbolic guidance, while the LLM supplies lexical selection, fluency, and generalization capabilities.

5.2. Mexican Sign Language Glosses to Spanish Prompt

The MSLG-to-Spanish prompt is designed as a linguistically guided reconstruction framework that transforms gloss-based symbolic representations into fluent and grammatically well-formed Spanish. Unlike the Spanish-to-MSLG prompt, which primarily focuses on reducing and reorganizing spoken-language structure into gloss-compatible sequences, the MSLG-to-Spanish prompt emphasizes the recovery of grammatical, temporal, and discourse information that is frequently implicit or underspecified in gloss annotations. The prompt therefore functions as an explicit interpretive procedure that guides the language model in expanding compressed gloss sequences into natural Spanish sentences.

A central design principle of the prompt is the treatment of glosses as partial semantic cues rather than as direct word-by-word equivalents. The prompt incorporates heuristics for reconstructing verbal tense, aspect, modality, argument structure, determiners, agreement, and syntactic relations that are often absent from gloss sequences. Particular attention is devoted to the interpretation of temporal framing and perfective meaning, including rules governing the omission or reinterpretation of markers such as YA when temporal reference is already established elsewhere in the sentence. Similarly, the prompt includes strategies for recovering implicit subjects, introducing appropriate function words, resolving possessive relations, and producing idiomatic Spanish constituent order.

The prompt also addresses several recurrent structural divergences between gloss representations and Spanish syntax. These include topic–comment organization, verb-chain interpretation, movement verbs, adjectival placement, clitic phenomena, support-verb constructions, and subordinate structures. Since gloss sequences frequently prioritize information structure over canonical spoken-language syntax, the prompt provides rules for reordering constituents according to Spanish morphosyntactic and pragmatic conventions. It also incorporates lexical-semantic heuristics intended to avoid excessively literal translations and to favor more idiomatic Spanish realizations when appropriate.

An important component of the design is the explicit handling of ambiguity and underspecification. Because glosses linearize only part of the multimodal information present in signing, many distinctions required in Spanish output must be inferred contextually by the model. The prompt therefore encourages interpretive reconstruction rather than strict symbolic substitution, allowing the LLM to exploit its broader linguistic knowledge while remaining constrained by explicit transfer instructions.

Conceptually, the prompt combines symbolic procedural guidance with probabilistic language generation. The prompt operates as an externalized interpretive manual layered over the latent multilingual competence of the language model. In this sense, the system resembles a hybrid architecture in which explicit linguistic heuristics constrain and shape the model's generative behavior, while the LLM supplies lexical fluency, contextual inference, and grammatical completion capabilities.

5.3. Pretrained Language Model

The experiments were initially conducted with the open-weight gpt-oss-20b language model released by OpenAI [16]. These models were selected because they offered strong multilingual capabilities, instruction-following behavior, and a suitable open-weight setting for prompting-based experimentation. Their reasoning-oriented variants were especially attractive for a translation framework based on procedural linguistic instructions and symbolic transfer heuristics. However, after the release of the Gemma 4 instruction-tuned models [17], we evaluated the 31B variant as an alternative backbone. Preliminary experiments showed that Gemma followed the prompt instructions more consistently than gpt-oss-20b, particularly with respect to output format, gloss conventions, constituent order, temporal and aspectual heuristics, and other explicit translation rules. Since the proposed system relies heavily on procedural prompting, this higher degree of instruction adherence became a decisive factor.

The final system therefore relies primarily on the Gemma instruction-tuned architecture. Gemma 4 models are decoder-only transformer language models designed for multilingual instruction following, reasoning, and efficient deployment in open-weight environments. The 31B-scale variant (gemma-4-31B-it) provided a favorable balance between multilingual competence, reasoning capacity, controllability, and feasibility for local inference on the available research infrastructure. In our experiments, gemma-4-31B-it produced outputs that more reliably respected gloss formatting conventions, temporal and aspectual heuristics, constituent-order constraints, and explicit lexical transfer instructions encoded in the prompts.

6. Experiments with the MSLG–SPA Corpus

Evaluation of generated text is typically carried out using both surface-based and semantic metrics. BLEU (Bilingual Evaluation Understudy) [18] is a widely used automatic metric that measures the overlap between a system's output and one or more reference texts in terms of n-gram precision. It provides a convenient and reproducible way to assess how closely a hypothesis matches the reference at the lexical and phrase level, but it is known to be sensitive to variation in wording and word order. This is a particularly important drawback in sign-language gloss translation, where several gloss sequences may be linguistically acceptable even if they do not match the reference exactly. BLEU can therefore penalize valid lexical alternatives, different but acceptable constituent orders, and outputs that preserve the intended meaning while using a different level of explicitness. Conversely, a relatively high BLEU score does not necessarily guarantee that the output is pragmatically adequate, structurally well formed, or faithful to the multimodal information that glosses only partially encode. For this reason, BLEU should be interpreted as an indicator of surface similarity rather than as a complete measure of translation quality.

To complement BLEU, we also employ SAS (Semantic Answer Similarity) [19], which estimates the semantic equivalence between the hypothesis and the reference using vector representations derived from pretrained language models. In our case, these representations are obtained with the Sentence-Transformers model `paraphrase-multilingual-mpnet-base-v2`, a multilingual sentence encoder from the Sentence-Transformers framework [20], designed to map semantically similar texts to nearby regions of the embedding space across languages. Unlike BLEU, SAS captures meaning beyond exact token matches and is therefore more robust to paraphrasing and structural variation. This is particularly useful in our setting, where two outputs may differ in wording or local syntactic realization while still conveying essentially the same content. However, SAS also has important limitations. Because it relies on sentence embeddings learned from large multilingual text corpora, it may overestimate adequacy when an output is semantically close at a broad sentence level but wrong in linguistically crucial details such as argument structure, temporal anchoring, negation or directionality. The underlying sentence-embedding model has been trained on multilingual textual data, including Spanish, but not on MSLG glosses as a linguistic system. For this reason, we apply SAS only in the MSLG-to-Spanish direction, where the target output is ordinary Spanish text. Applying the metric directly to symbolic MSLG sequences would risk treating glosses as ordinary sentences and would not reliably capture the

Table 1
Top system runs by BLEU score for the Spanish-to-MSLG direction in the MSLG–SPA shared task at IberLEF 2026

Rank	System	BLEU
1	AxoloTux_C_SPA2MSLG	26.4909
2	AxoloTux_QDistilTest_SPA2MSLG	24.4209
3	VerbaNexAI_ManyShot_SPA2MSLG	23.7883
4	AxoloTux_QDistil2_SPA2MSLG	23.2397
5	VerbaNexAI_FewShot10RagCurriculum_SPA2MSLG	22.3941
6	UAM_ChineseRoom_SPA2MSLG	21.3552
7	CICESE-LCDAA_run4_SPA2MSLG	21.1909
8	CICESE-LCDAA_run0_SPA2MSLG	20.2347
9	AxoloTux_QBase_SPA2MSLG	19.2774
10	CICESE-LCDAA_run1_SPA2MSLG	18.9380

linguistic significance of gloss conventions or sign-language-specific ordering. SAS should therefore be interpreted as an approximate semantic similarity measure, not as a substitute for linguistic evaluation. Together, BLEU and SAS offer a more comprehensive evaluation, combining surface fidelity with semantic adequacy.

The following subsections report on evaluation and examines representative outputs obtained with the shared-task data in order to identify the main strengths and limitations of the prompting strategy in both translation directions. Rather than focusing only on overall adequacy, the analysis considers how the system handles recurrent linguistic phenomena such as temporal framing, constituent order, argument structure, gloss selection, referential explicitness, and regionally marked Spanish expressions.

6.1. Spanish to Mexican Sign Language Glosses

Table 1 shows that the Spanish-to-MSLG direction is considerably more difficult than the reverse direction. The best-ranked run reaches 26.49 BLEU, and most of the top systems cluster in a relatively narrow band between roughly 19 and 26 BLEU. Our system run, UAM_ChineseRoom, ranks sixth with 21.35 BLEU, placing it 5.13 BLEU points behind the leading system but still within the main competitive range of submitted systems. This pattern suggests that the task is not primarily separated by a single dominant approach; rather, several systems achieve comparable mid-range performance while facing similar difficulties in generating compact and correctly ordered MSLG sequences from morphologically richer Spanish input.

Additional experiments carried out on the training set during development should be interpreted with caution, since a small number of training examples were also used as in-context examples in the prompt. For this reason, the training scores are best understood as an internal sanity check on the behavior of the prompting strategy, rather than as an independent estimate of generalization performance. Under this setting, the SPA-to-MSLG system obtains a BLEU score of 29.91, which points to moderate performance. This score suggests that the model often produces understandable gloss sequences and recovers many appropriate lexical items, but still falls short of consistently high-quality translation. In practical terms, the output is useful for analysis, although noticeable errors remain, especially in fluency, ordering, and structural accuracy. A closer examination of the n-gram precisions reveals a clear pattern. Unigram precision is relatively high (78.1%), indicating that the model successfully selects appropriate words or signs in many cases. However, performance drops sharply as the n-gram order increases: 36.5% for bigrams, 21.1% for trigrams, and 13.8% for four-grams. This decline suggests that, while the system captures local lexical choices reasonably well, it struggles to reproduce longer token sequences. In other words, the main limitation lies not only in vocabulary selection, but also in arranging and combining elements into well-formed MSLG structures.

Overall, the results point to a system that captures content but fails to fully model the target language's syntactic and phrase-level organization. Improvements should therefore focus on better handling of word order, phrase structure, and multi-token dependencies. Addressing these aspects is likely to yield

significant gains in higher-order n-gram precision and, consequently, in overall BLEU performance.

We have not applied SAS in this translation direction because the output consists of gloss sequences, whose evaluation depends not only on semantic overlap, but also on conformity with annotation conventions and on the structural decisions used to linearize signed content. Two gloss sequences may be semantically close while differing in temporal framing, constituent order, lexical choice, compounding, or the explicit marking of discourse and referential information. In such cases, a semantic-similarity measure could overestimate performance by overlooking differences that are important for the target gloss representation. For this reason, BLEU and qualitative error analysis are more informative here than SAS alone.

The following examples provide a qualitative view of these issues. They were selected to illustrate both successful behavior and recurrent error patterns in the Spanish-to-MSLG direction, including temporal anchoring, topicalization, regional lexical interpretation, argument structure, dependency relations, referential explicitness, and pragmatic compression.

MSLG–SPA corpus example 312, shown in (1), illustrates the model's effective handling of temporal framing, contrast, and Mexican Spanish nuance. In this case, the model interprets *corriente* as 'low-quality/cheap' rather than simply 'ordinary'. It places AYER ('yesterday') in sentence-initial position, topicalizes ESTE CELULAR ('this cellphone'), adds the explicit subject YO ('I'), marks event completion with YA ('already'), and preserves the adversative relation with PERO ('but').

(1) a. *Compré este celular ayer pero está muy corriente*
'I bought this cellphone yesterday, but it is very cheap/low-quality'

b. AYER ESTE CELULAR YO YA COMPRAR PERO CORRIENTE
yesterday this cellphone I already buy but low-quality

Example 702, shown in (2), provides a straightforward case in which redundant YA ('already') is avoided because AÑO-PASADO ('last year') already establishes past time.

(2) a. *Mi generación se graduó el año pasado*
My class graduated last year

b. AÑO-PASADO MI GENERACIÓN GRADUAR
last-year my class graduate

Example 804, shown in (3), highlights the failure in the reordering of a complex verbal chain. The Spanish sequence combines obligation (*tengo que*), habituation (*acostumbrarme*), waking (*despertarme*), and purpose (*correr*). The model (3c) reorganizes these verbal relations as CORRER ('to run'), DESPERTAR ('wake up'), ACOSTUMBRAR ('get used to'), and DEBER ('must'), while fronting the temporal frame TEMPRANO ('early').

(3) a. *Tengo que acostumbrarme a despertarme a correr temprano*
'I have to get used to waking up to run early'

b. TEMPRANO DESPERTAR CORRER YO DEBER COSTUMBRE
early wake-up run I must get-used-to

c. TEMPRANO CORRER DESPERTAR ACOSTUMBRAR YO DEBER
early run wake-up get-used-to I must

Example 634, shown in (4), illustrates how the system handles an exhortative first-person plural construction. The Spanish verb *caminemos* encodes both the imperative/exhortative force ('let us walk') and an implicit first-person plural subject, which the model makes explicit as NOSOTROS ('we'). It also preserves the motion predicate CAMINAR ('walk'), renders the directional phrase *hacia el sur* as A SUR (lit. 'to south'), and retains PARA ('for') to mark the purpose relation linking AGUA ('water') and ENCONTRAR ('find').

(4) a. *Caminemos hacia el sur para encontrar agua*
'Let us walk south to find water'

b. NOSOTROS CAMINAR A SUR PARA AGUA ENCONTRAR
we walk to south for water find

Example 875, shown in (5), illustrates the system's handling of topic–comment structure and its divergence from the reference ordering. Both solutions avoid a literal rendering of the Spanish preposition *en* ('in'), since the belief relation is expressed directly by the predicate CREER ('believe'). However, they organize the remaining material differently. The reference translation places CRISTIANA in initial position and leaves FE at the end of the sequence, whereas the system keeps the nominal expression FE CRISTIANA ('Christian faith') together and fronts it as the topic before the comment YO CREER ('I believe'). The system output in (5c) is therefore semantically transparent and compatible with a topic–comment analysis, but it diverges from the specific word order encoded in the reference annotation in (5b).

(5) a. *Yo creo en la fe cristiana*
'I believe in the Christian faith'

b. CRISTIANA YO CREER FE
Christian I believe faith

c. FE CRISTIANA YO CREER
faith Christian I believe

The reasoning trace below makes explicit why the system arrives at this ordering. It first identifies the main semantic roles in the Spanish sentence: *yo* as the experiencer or subject of belief, *creo* as the predicate, and *la fe cristiana* as the object of belief. It then removes Spanish-specific material that has no direct gloss counterpart, such as the article *la* and the preposition *en*, and maps the remaining content words to YO, CREER, FE, and CRISTIANA. The crucial step is the choice between two competing ordering principles. A more strictly transitive template would produce a subject–object–verb order, whereas the topic–comment heuristic favors placing the object of belief first. The system resolves this conflict by treating FE CRISTIANA as the discourse topic and YO CREER as the comment. This explains why the system output differs from the reference while still reflecting a linguistically motivated analysis of the sentence.

```
*   Input: "Yo creo en la fe cristiana."
*   Meaning: "I believe in the Christian faith."

*   Temporal frame: None.
*   Locative frame: None.
*   Participants: Yo (Subject).
*   Main event: creer (Predicate).
*   Object/Topic: la fe cristiana (Object of belief).
*   Polarity: Positive.
*   Modality: Declarative.
*   Sentence type: Statement.

*   Subject: "Yo" -> YO.
*   Verb: "creo" -> CREER.
*   Object: "la fe cristiana" -> FE CRISTIANA.
*   Preposition: "en" (usually omitted or implied in belief contexts in MSLG).

*   C7 (Transitive verbs): [SUBJECT] [OBJECT] [VERB].
*   C3 (Topic-comment): [TOPIC] [SUBJ] [PRED].
*   In MSLG, for beliefs or convictions, the object of the belief often acts as the
↪  topic.
*   Structure: FE CRISTIANA YO CREER.
```

```
*   Uppercase: FE CRISTIANA YO CREER.
*   Remove articles: "la" is gone.
*   Check for proper names: "cristiana" is an adjective here, not a proper name of a
↪  person.
*   Check for compounds: "fe cristiana" is a standard noun-adjective pair.

*   Input: Yo creo en la fe cristiana.
*   Subject: YO
*   Object: FE CRISTIANA
*   Verb: CREER
*   Order: FE CRISTIANA YO CREER.

*Wait, let's check if "creer" is treated as a transitive verb or if the subject
↪  should come first.*
C7 says [SUBJECT] [OBJECT] [VERB].
C3 says prefer topic-comment.
"Fe cristiana" is the topic.
Result: FE CRISTIANA YO CREER.

*Alternative check:*
If I follow C7 strictly: YO FE CRISTIANA CREER.
However, in MSLG, the object of belief is typically topicalized.
Let's look at the few-shots: "Necesito un cuaderno" -> "CUADERNO YO NECESITAR".
↪  This supports the Topic-Subject-Verb pattern.

Final sequence: FE CRISTIANA YO CREER.
```

The trace therefore shows that the system's decision is not a random ordering error, but the outcome of a rule conflict resolved in favor of topicalization. The remaining discrepancy lies in the fact that the reference annotation separates CRISTIANA and FE, while the system keeps the nominal expression together as a single topical unit.

Example 652, shown in (6), illustrates the treatment of clitic doubling and experiencer structure. The Spanish construction *a mí no me gustan* contains both a prepositional dative phrase, *a mí*, and the dative clitic *me*. These two elements do not introduce two separate participants; rather, they jointly mark the experiencer of *gustar*. The system correctly reconstructs this experiencer as YO ('I') and encodes the negative evaluative predicate with the lexicalized sign NO-GUSTAR ('not-like'). It also reduces the plural noun *ametralladoras* ('machine guns') to the singular base gloss AMETRALLADORA ('machine gun'), a common glossing strategy when number is not overtly marked. The main difference between the reference and the system output is the presence of DIFERENTE ('different') in the reference. This additional gloss appears to strengthen or specify the contrastive/evaluative force of the utterance, whereas the system produces the more compact sequence AMETRALLADORA YO NO-GUSTAR. The system output is therefore semantically adequate and structurally economical, but it omits an extra evaluative or contrastive element encoded in the reference annotation.

(6) a. *A mí no me gustan las ametralladoras*
'I do not like machine guns'

b. AMETRALLADORA DIFERENTE YO NO-GUSTAR
machine-gun different I not-like

c. AMETRALLADORA YO NO-GUSTAR
machine-gun I not-like

6.2. Mexican Sign Language Glosses to Spanish

Table 2 shows that the MSLG-to-Spanish direction is highly competitive across the top-ranked systems in the MSLG–SPA shared task at IberLEF 2026. The best submission reaches a BLEU score of 54.03,

Table 2
Top systems by BLEU score for the MSLG-to-Spanish direction in the MSLG–SPA shared task at IberLEF 2026

Rank	System	BLEU
1	CICESE-LCDAA_run0_MSLG2SPA	54.0308
2	AxoloTux_C_MSLG2SPA	52.7022
3	AxoloTux_QDistilTest_MSLG2SPA	51.1455
4	UAM_ChineseRoom_MSLG2SPA	50.5257
5	UCOUCPlenitas_mistral_MSLG2SPA	47,6182
6	UCOUCPlenitas_qwen_MSLG2SPA	47,6182
7	AxoloTux_G_MSLG2SPA	47,0954
8	UCOUCPlenitas_gemma4_cloud_MSLG2SPA	46,9379
9	AxoloTux_QDistilMix_MSLG2SPA	46,7560
10	Ludwig_QW-POS75DA_MSL2SPA	45,4934

while our system (UAM_ChineseRoom) ranks fourth with 50.52 BLEU, only 3.50 BLEU points behind the first-ranked system. This relatively narrow gap suggests that the prompting-based approach remains competitive in this translation direction, even without task-specific fine-tuning.

During system development, the training-set evaluation yields a BLEU score of 50.16, indicating a strong degree of correspondence with the reference translations. This result suggests that the model often produces fluent and accurate Spanish outputs that remain close to the target formulations. The n-gram precision breakdown provides a more detailed view of this behavior: unigram precision is high (77.4%), showing that the system usually selects appropriate lexical items. Higher-order n-gram scores also remain comparatively strong, with 57.4% for bigrams, 44.4% for trigrams, and 34.9% for four-grams. The gradual decrease across n-gram levels suggests that the system captures not only individual lexical choices, but also local sequencing, phrase structure, and short-range syntactic dependencies. Overall, these results characterize a system that performs well both lexically and structurally. The relatively strong higher-order n-gram scores indicate that the model is able to generate coherent Spanish sequences and generally respect target-language ordering patterns. Remaining improvements are likely to involve subtle aspects of fluency, completeness, and context-sensitive morphosyntactic reconstruction, but the system already shows a strong capacity to produce accurate and well-structured outputs.

The SAS metric further supports this interpretation. With a mean of 0.944, the system shows very strong overall semantic alignment with the reference translations. The maximum value of 1.000 indicates that many outputs achieve complete semantic equivalence, whereas the minimum value of 0.320 points to a small number of substantial errors or mismatches. Taken together, these results suggest that the system generally preserves meaning very effectively, although occasional failures remain and should not be overlooked.

The following selected examples complement the automatic scores with a qualitative analysis of both successful reconstructions and remaining failure cases. They show that the system can often recover implicit Spanish morphosyntax, participant structure, regional lexical choices, and idiomatic expressions from compressed gloss sequences. At the same time, they also reveal where the approach remains fragile, especially in cases involving variety-specific lexical meanings, subtle copula selection, tense interpretation, and pragmatic distinctions that are only partially encoded in the gloss representation.

MSLG–SPA training corpus example 220, shown in (7), illustrates grammatical reconstruction from a highly compressed gloss sequence. The model infers plural number in the noun phrase and predicate (*peces*, *azules*, *favoritos*), introduces the definite article (*los*), reconstructs the predicative copula (*son*), and reorganizes the sequence into a fully grammatical Spanish sentence. This example suggests that the system is not merely performing literal lexical substitution, but also recovering latent morphosyntactic structure absent from the gloss representation itself.

(7) a. PEZ COLOR AZUL MI FAVORITO
fish color blue my favorite
b. Los peces azules son mis favoritos

'Blue fish are my favorites'

Example 768, shown in (8), provides a clear case of participant reconstruction and regional lexical selection. In the system output, the model interprets DOS-DE-NOSOTROS as a dual participant configuration corresponding to *tú y yo* in (8c), reconstructs the reciprocal verbal structure *nos vemos*, and selects the Mexican Spanish temporal expression *en la noche* rather than the more common European Spanish alternative *por la noche*.

(8) a. DOS-DE-NOSOTROS NOS-VEMOS MAÑANA NOCHE
two-of-us see-each-other tomorrow night

b. *Nos vemos mañana en la noche*
'See you tomorrow night'

c. *Tú y yo nos vemos mañana en la noche*
'You and I will see each other tomorrow night'

Example 402, shown in (9), illustrates successful lexical and gender reconstruction despite an error in tense interpretation. The gloss PONER-ATENCIÓN ('pay-attention'), together with MAESTRO+MUJER ('teacher+woman'), leads the model to produce *maestra* and the idiomatic Mexican Spanish expression *poner atención*. This example indicates that the system can recover feminine gender and idiomatic lexical choices even when other aspects of the sentence remain imperfect.

(9) a. ¿TUYA MAESTRO+MUJER TÚ PONER-ATENCIÓN ELLA?
your teacher+woman you pay-attention she

b. ¿Le pusiste atención a tu maestra?
Did you pay attention to your teacher?

c. ¿Le pones atención a tu maestra?
Do you pay attention to your teacher?

Example 617, shown in (10), illustrates a clear lexical-semantic failure involving a domain-specific gloss. In this context, ARCHIVO should be interpreted through the Mexican Spanish administrative verb *capturar* ('to enter' or 'to type in' data), as in entering information into a computer. The model instead selects the more general lexical equivalent *archivar* ('to file' or 'to archive') and therefore produces a fluent but semantically incorrect sentence. The case shows that plausible Spanish output does not guarantee correct interpretation of domain- and variety-specific lexical mappings.

(10) a. SECRETARIA ELLA COMPUTADORA CARTA ARCHIVO
secretary she computer letter archive/file
'The secretary enters/types the letter into the computer'

b. *La secretaria está capturando la carta*
'The secretary is entering the letter into the computer'

c. *La secretaria archiva la carta en la computadora*
'The secretary archives the letter on the computer'

Example 659, shown in (11), illustrates a pragmatically relevant choice between the Spanish copulas *estar* and *ser*. The gloss sequence ESTE JUGUETE BARATO (lit. 'this toy cheap') is better rendered as *Este juguete está barato* ('This toy is cheap right now'), because Mexican Spanish commonly uses *estar barato* to refer to the current price or temporary affordability of a specific item. By contrast, *ser barato* ('to be cheap' as an inherent or generic property) remains grammatically acceptable, but it weakens the intended contextual reading.

(11) a. ESTE JUGUETE BARATO
this toy cheap

b. *Este juguete está barato*
   'This toy is cheap (right now)'
c. *Este juguete es barato*
   'This toy is cheap'

The system's reasoning trace is informative because it shows where the error enters the reconstruction process. The model correctly treats the gloss sequence as a predicative structure rather than as a noun phrase, and therefore inserts an implicit Spanish copula. However, it then selects *ser* rather than the pragmatically more appropriate *estar* for a context-dependent price evaluation in Mexican Spanish:

```
Input: 'ESTE JUGUETE BARATO'

Glosses:
  * 'ESTE': Demonstrative (this)
  * 'JUGUETE': Noun (toy)
  * 'BARATO': Adjective (cheap)

A1: Do not translate literally.
A2: Reorder for natural Spanish.
A3: Add articles/copulas if necessary.
A16: Do not invent missing information.

Literal: "Este juguete barato"

Possible interpretations:
  1.  A noun phrase: "Este juguete barato" (This cheap toy).
  2.  A sentence with an implicit copula: "Este juguete es barato" (This toy is
  ↪  cheap).

* In MSLG, a simple sequence of [Subject] [Adjective] often functions as a sentence
↪  "The [Subject] is [Adjective]".
* However, without a clear predicate or context, it could be just a phrase.
* Comparing with examples: 'PEZ COLOR AZUL MI FAVORITO' -> 'Los peces azules son mis
↪  favoritos.' (Copula added).
* 'MI HERMANO+MUJER YA EMBARAZADA' -> 'Mi hermana está embarazada.' (Copula added).
* Therefore, it's more likely to be a statement: "Este juguete es barato."

Output: "Este juguete es barato." (Natural Mexican Spanish).
```

7. Limitations

Because the corpus consists primarily of isolated sentence-level pairs, information-structural distinctions such as topic, focus, contrast, and discourse prominence are often only partially recoverable. This limitation is particularly relevant for gloss-based representations, where non-manual and discourse-level cues associated with topic–comment structure may be absent or underspecified. At the same time, this issue is central to sign language translation, since constituent order in sign languages is frequently shaped by information structure rather than by a fixed canonical word order alone. Consequently, gloss sequences may reflect topic–comment organization, discourse prominence, or visual-spatial reference strategies that cannot always be reconstructed unambiguously from isolated examples.

A second limitation concerns the size and coverage of the corpus. With only a few hundred aligned training pairs, the dataset cannot represent the full lexical, syntactic, and discourse diversity of either Mexican Sign Language or Mexican Spanish. Rare constructions, low-frequency lexical items, regional variants, and pragmatically marked expressions are therefore likely to be underrepresented, which restricts both the generalizability of the prompts and the robustness of the resulting translations.

A further limitation is the dependence of the approach on manually designed prompt heuristics. While this makes the system more interpretable, it also means that performance is sensitive to the quality, completeness, and transferability of the rules encoded in the prompts. Heuristics that work well for recurrent patterns in this corpus may not extend reliably to new domains, longer discourse contexts, or different annotation conventions.

Finally, the evaluation remains constrained by the nature of the task and the available references. In gloss-based translation, exact or near-exact match with a reference can obscure the fact that multiple outputs may be semantically adequate while differing in word order, lexical choice, or level of explicitness. The reported results should therefore be interpreted as informative but partial indicators of translation quality rather than as a complete measure of linguistic adequacy.

8. Conclusions

The rankings in Tables 1 and 2 highlight the asymmetry between the two translation directions. Whereas MSLG-to-Spanish translation can benefit from the language model's strong prior over fluent Spanish, Spanish-to-MSLG requires the system to suppress overt Spanish morphology, select appropriate gloss labels, and reorganize the sentence into the more compressed conventions of the target gloss representation. The lower BLEU scores in Table 1 therefore reflect not only lexical errors, but also the structural difficulty of producing the expected gloss order and level of explicitness.

The results of this work suggest that prompting can provide a viable strategy for bidirectional translation between Spanish and Mexican Sign Language glosses in low-resource conditions. Across both translation directions, the system is often able to reconstruct information that is absent or only partially explicit in the input, including temporal framing, argument structure, lexical composition, referential relations, and regionally marked Spanish choices. At the same time, the error analysis shows that these strengths remain uneven: failures in gloss selection, event-structure interpretation, pragmatic lexical choice, and hierarchical dependency recovery reveal the limits of operating over compressed symbolic representations. Taken together, these results show that carefully designed prompts can serve as an interpretable method for low-resource sign-language translation, while also exposing the conceptual boundaries of that achievement. The system works precisely because it exploits structured symbolic proxies, and therefore inherits their representational limits.

These findings are especially significant in light of the Chinese Room analogy that motivates the paper. In the present system, the prompt functions as an external procedural manual for segmenting, reordering, normalizing, and reconstructing symbolic inputs. At the same time, the system is not simply a rule-based translator. Its behavior also depends on the statistical regularities encoded in the pretrained language model. It is therefore best understood as a stochastic parrot operating inside Searle's Chinese Room: the prompt supplies the rule book, the language model supplies distributional fluency, and the gloss representation supplies the symbolic interface. This combination enables useful transformations between written Spanish and Mexican Sign Language glosses, but it should not be treated as evidence of grounded understanding.

9. Future Work

Future work could incorporate more explicit insights from research on constituent order in sign languages. As Napoli and Sutton-Spence [21] argue, word order in sign languages is not arbitrary; rather, it emerges from the interaction between general linguistic pressures and modality-specific constraints associated with visual communication. Their survey of 42 sign languages shows that sign languages overwhelmingly favor SOV (Subject-Object-Verb) or SVO patterns, while also exhibiting substantial flexibility driven by discourse structure and visual-spatial organization. Future prompts should therefore incorporate information such as the following generalizations when translating into sign language glosses:

1. SOV order is grammatical in all sign languages surveyed.

2. If an argument affects the phonological shape of the verb, it tends to precede the verb. This includes classifier predicates, agreeing verbs, pointing verbs, spatial verbs and argument-sensitive verbs.
3. The most common sentence type contains only one new argument, and that argument tends to precede the verb.
4. When two manually expressed noun phrases occur in a locational expression within a single clause, larger or less mobile referents tend to precede smaller or more mobile referents, regardless of theta role or grammatical function.
5. The object tends to be immediately adjacent to the verb.
6. In reversible sentences with plain verbs, SVO order is favored.

These observations have direct consequences for prompting: they should not remain only as background linguistic observations, but could be encoded as explicit procedural guidance for Spanish-to-MSLG translation and other pairs or representations.

Declaration on Generative AI

During the preparation of this work, the author used GPT-5.5 and Prism from OpenAI for text translation, paraphrasing and rewording, improving writing style, drafting the abstract, grammar and spelling checks, and peer-review simulation. After using these tools and services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] A. Moryossef, Y. Goldberg, Sign Language Processing, https://sign-language-processing.github.io/, 2021.

[2] V. Sutton, Lessons in SignWriting Textbook, The SignWriting Press, La Jolla, CA, 2002. Available online from the Center for Sutton Movement Writing.

[3] T. Hanke, HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts, in: Proceedings of the LREC2004 Workshop on the Representation and Processing of Sign Languages: From SignWriting to Image Processing, European Language Resources Association (ELRA), Lisbon, Portugal, 2004, pp. 1–6.

[4] 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).

[5] 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.

[6] J. R. Searle, Minds, brains, and programs, Behavioral and Brain Sciences 3 (1980) 417–457. doi:10.1017/S0140525X00005756.

[7] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 610–623. doi:10.1145/3442188.3445922.

[8] M. Cruz-Aldrete, Gramática de la Lengua de Señas Mexicana, Tesis doctoral, Centro de Estudios Lingüísticos y Literarios, El Colegio de México, 2008.

[9] W. Sandler, D. Lillo-Martin, Sign Language and Linguistic Universals, Cambridge University Press, 2006. doi:10.1017/CBO9781139163910.

[10] J. Porta, F. López-Colino, J. Tejedor, J. Colás, A rule-based translation from written spanish to spanish sign language glosses, Computer Speech Language 28 (2014) 788–811. doi:10.1016/j.csl.2013.10.003.

[11] R. San Segundo Hernández, J. M. Montero Martínez, R. d. Córdoba Herralde, V. Sama, F. Fernández Martínez, L. F. D'Haro Enríquez, 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. doi:10.1007/s10044-011-0243-9.

[12] L. Chiruzzo, E. McGill, S. Egea-Gómez, H. Saggion, Translating Spanish into Spanish Sign Language: Combining rules and data-driven approaches, in: A. K. Ojha, C.-H. Liu, E. Vylomova, J. Abbott, J. Washington, N. Oco, T. A. Pirinen, V. Malykh, V. Logacheva, X. Zhao (Eds.), Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022), Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 75–83. URL: https://aclanthology.org/2022.loresmt-1.10/.

[13] S. Yazdani, J. V. Genabith, C. España-Bonet, Continual learning in multilingual sign language translation, in: L. Chiruzzo, A. Ritter, L. Wang (Eds.), Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Albuquerque, New Mexico, 2025, pp. 10923–10938. doi:10.18653/v1/2025.naacl-long.546.

[14] A. Núñez-Marcos, O. P. de Viñaspre, G. Labaka, A survey on Sign Language machine translation, Expert Systems with Applications 213 (2023) 118993. doi:https://doi.org/10.1016/j.eswa.2022.118993.

[15] G. Tanzer, M. Suzgun, E. Visser, D. Jurafsky, L. Melas-Kyriazi, A Benchmark for Learning to Translate a New Language from One Grammar Book, in: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024, OpenReview.net, 2024. URL: https://openreview.net/forum?id=tbVWug9f2h.

[16] S. Agarwal, L. Ahmad, J. Ai, S. Altman, et al., gpt-oss-120b & gpt-oss-20b model card, 2025. URL: https://arxiv.org/abs/2508.10925.

[17] Google DeepMind, Gemma 4 model card, 2026. URL: https://ai.google.dev/gemma/docs/core/model_card_4.

[18] 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, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 2002, pp. 311–318. doi:10.3115/1073083.1073135.

[19] J. Risch, T. Möller, J. Gutsch, M. Pietsch, Semantic Answer Similarity for Evaluating Question Answering Models, in: A. Fisch, A. Talmor, D. Chen, E. Choi, M. Seo, P. Lewis, R. Jia, S. Min (Eds.), Proceedings of the 3rd Workshop on Machine Reading for Question Answering, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 149–157. doi:10.18653/v1/2021.mrqa-1.15.

[20] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. doi:10.18653/v1/D19-1410.

[21] D. J. Napoli, R. Sutton-Spence, Order of the major constituents in sign languages: Implications for all language, Frontiers in Psychology 5 (2014) 376. doi:10.3389/fpsyg.2014.00376.

A. SPA-to-MSLG Prompt

```
You are an expert translator from Mexican Spanish into MSLG (Mexican Sign Language
↪ glosses)

Your task is to translate a sentence in Mexican Spanish into a sequence of MSL
↪ glosses.
```

```
# ==================== 1. OBJECTIVE ====================

Input:

- A sentence or short text in Spanish.

Output:

- A single sequence of linear glosses in uppercase, preserving accent marks and
↪  diaereses.

Output format:

- glosses separated by spaces
- preserve ¿? and ¡! when applicable
- put `dm-` before anthroponyms
- do not put `dm-` before toponyms
- put `#` before initials and acronyms
- use hyphens in compound glosses when appropriate
- no comments, no prefixes, no notes

# ==================== 2. PIPELINE ====================

ALWAYS process in these four phases:

PHASE A. PREPROCESSING
PHASE B. LEXICO-SYNTACTIC TRANSFER
PHASE C. STRUCTURAL REORDERING
PHASE D. POSTPROCESSING AND NORMALIZATION

# ==================== 3. PHASE A. PREPROCESSING ====================

A1. Segment the input and identify:

- temporal frame
- locative frame
- participants
- main event
- polarity
- modality
- sentence type

A2. Classify the constituents using these internal labels:

- TIME
- LOCATION
- TOPIC
- SUBJECT
- OBJECT
- PREDICATE
- MODAL
- ASPECT
- NEG
- WH
- EVALUATION

A3. Detect explicit signals in Spanish:

- lexical time: ayer, hoy, mañana, domingo, el año pasado, próximo año...
```

- place: en X, a X, desde X...
- negation: no, nunca, nadie, nada, sin...
- modality: deber, poder, querer, ojalá, tener que...
- aspect: ya, todavía, apenas, seguir, acabar de...
- question: por qué, dónde, quién, cuál, cuántos, qué...
- proper names: anthroponyms, toponyms, brands, organizations, institutions, events, etc.
- possessive relations
- attributive adjectives
- verbs of motion, perception, state, possession, and existence

A4. Reduce the input to a minimal semantic representation:

- keep only the main content
- remove Spanish grammatical function that is not needed in the final gloss

==================== 4. PHASE B. LEXICO-SYNTACTIC TRANSFER ====================

B1. Subjects

Spanish frequently omits subject pronouns because person and number are encoded in verbal morphology. In MSLG, however, the subject is often expressed explicitly to maintain clarity and discourse coherence, especially when the subject is not recoverable from the immediate context.

Therefore, when translating from Spanish into MSLG:

- If the subject is omitted in Spanish but can be clearly inferred, explicitly introduce the corresponding pronoun or referent in MSLG when needed for disambiguation or fluency.

- Explicit subjects are especially common with:

    * first and second person predicates,
    * verb chains,
    * changes of subject,
    * coordinated or subordinate clauses,
    * and contexts where omission could create ambiguity.

Examples:

- Quiero vivir -> YO QUERER VIVIR
- Mañana iremos al mercado -> MAÑANA NOSOTROS MERCADO IR
- Dice que vendrá -> ÉL DECIR QUE VENIR

B2. Verbs

Convert conjugated verbs into base lexical glosses:

- llegó -> LLEGAR
- compré -> COMPRAR
- vive -> VIVIR
- conociste -> CONOCER

B3. Tense, aspect, and modality

Do not use Spanish verbal inflection; use lexical particles:

- perfective / completed -> YA

- not completed -> TODAVÍA
- future -> MAÑANA / PRÓXIMO / FUTURO
- obligation -> DEBER
- possibility -> PODER
- desire -> QUERER / OJALÁ
- immediacy -> AHORITA / AHORA
- anteriority -> ANTES / AÑO-PASADO / HACE-PASADO

YA tends to go before the verb as a marker of a past or completed event, except when it closes the event, in which case it goes at the end. When there is an explicit temporal adverb, YA is not necessary.

B4. Negation

Map negation to the most natural pattern:

- NO
- NO-HAY
- NO-GUSTAR
- NINGÚN
- NO-PODER
- NO-FUI

Do not force a single strategy for all cases.

B5. Copula and attributive predication

If Spanish uses SER/ESTAR + attribute, normally remove the copula.

Examples:

- es barato -> BARATO
- está caliente -> CALIENTE
- es famosa -> FAMOSO

B6. Existence and possession

- haber / hay -> HABER
- no haber / no hay -> NO-HAY
- tener -> TENER or a possessive/topicalized structure
- location -> locative order or simple nominal sequence, without copula if possible

B7. The verb "tener"

When the verb "tener" is used as a light verb with an abstract mass noun denoting a property or quality as direct object, gloss it with the equivalent verb or noun.

Examples:

- Ulises tiene dolor en la rodilla -> dm-ULISES RODILLA ÉL DOLER
- Manuel tiene cuidado -> dm-MANUEL CUIDAR
- Jordi tiene sed -> dm-JORDI SED
- Alejandro tuvo mucho miedo -> dm-ALEJANDRO MUCHO MIEDO

When the verb "tener" expresses age, it may be omitted, although the noun EDAD is glossed.

Example:

- María tiene 18 años -> dm-MARÍA 18 EDAD

B8. Pronouns and possessives

Make them explicit when they provide structural clarity or when they are subjects or objects:

- YO, TÚ, ÉL, ELLA, NOSOTROS-DE-DOS
- MI, TUYO, SU, SUYO

Do not add unnecessary pronouns.

B9. Proper names

Add `dm-` before anthroponymic names.

Examples:

- Isabel -> dm-ISABEL
- Juana -> dm-JUANA
- Edgar -> dm-EDGAR

Do not use `dm-` for names of deities, saints, political or historical figures, etc.

Examples:

- Dios -> DIOS
- Adán -> ADÁN
- Jehová -> JEHOVÁ
- Jesús -> JESÚS
- Cristóbal Colón -> CRISTOBAL-COLÓN
- Charles-Michel L´Épée -> CHARLES-MICHEL-L´ÉPÉE
- Espíritu Santo -> ESPÍRITU-SANTO
- Adolfo López Mateo -> ADOLFO-LÓPEZ-MATEO (presidente de México)
- Enrique Peña Nieto -> ENRIQUE-PEÑA-NIETO (gobernador)

Do not add `dm-` before toponyms.

Examples:

- México -> MÉXICO
- Taxqueña -> TAXQUEÑA
- Europa -> EUROPA
- Islandia -> ISLANDIA
- Asia -> ASIA
- Filipinas -> FILIPINAS
- Estado-de-México -> ESTADO-DE-MÉXICO
- Villahermosa -> VILLAHERMOSA
- Ameria -> AMERIA
- Austria -> AUSTRÍA
- Ciudad Juárez -> CIUDAD-JUÁREZ
- Bangladesh -> BANGLADESH

Do not add `dm-` before organizations, institutions, brands, companies, works, etc.

Examples:

- Cruz-Roja -> CRUZ-ROJA
- Gamesa (marca de galletas) -> GAMESA
- Comercial Mexicana -> COMERCIAL MEXICANA

- Universidad Gallaudet -> UNIVERSIDAD GALLAUDET
- Plaza Universidad -> PLAZA UNIVERSIDAD
- Corán -> CORÁN
- Cepamex -> CEPAMEX
- Sabritas (fábrica) -> SABRITAS
- Palacio de los Deportes -> PALACIO-DE-LOS-DEPORTES
- Deportivo Galeana -> DEPORTIVO-GALEANA
- Teletón -> TELETÓN
- Wal-Mart -> WAL-MART
- Alcohólicos Anónimos -> ALCOHÓLICOS-ANÓNIMOS

Add `#` before initials and acronyms except LSM (lengua de signos mexicana) and CD (compact disc):

Examples:

- TV (televisión) -> #TV
- SEP (Secretaría de Educación Pública) -> #SEP
- VGT (Lengua de Signos Flamenca) -> #VGT
- LSFB (Lengua de Signos Franco-Belga) -> #LSFB

B10. Adjectives

Place the adjective after the noun:

Examples:

- nuevo coche -> COCHE NUEVO
- isla bonita -> ISLA BONITO

Color adjectives take the gloss COLOR before them:

Examples:

- rojo -> COLOR ROJO
- rojos -> COLOR ROJO
- verde y rojo -> COLOR VERDE ROJO

B11. Medical professions

Medical professions are compounds that use the gloss DOCTOR followed by the organ they treat:

Examples:

- cardiólogo ->DOCTOR-CORAZÓN
- oculista -> DOCTOR-OJOS
- ginecólogo -> DOCTOR-MUJER
- traumatólogo -> DOCTOR-HUESOS
- odontólogo -> DOCTOR-DIENTES
- ortopedista -> DOCTOR-HUESOS

B12. Compounds or idiomatic expressions:

Form compounds such as the following:

- bolsa de mano -> BOLSA-MANO
- crema para la cara -> CREMA-DE-COSMÉTICO
- licencia de conducir -> LICENCIA-DE-CONDUCIR

- control de videojuego -> CONTROL-DE-VIDEOJUEGO
- cinturón de seguridad -> CINTURÓN-DE-SEGURIDAD
- no hay -> NO-HAY
- lago de Pátzcuaro -> LAGO-DE-PÁTZCUARO
- carretilla de carga -> CARRETILLA-DE-CARGA
- por favor -> POR-FAVOR
- coca cola -> COCA-COLA
- baby shower -> BABY-SHOWER
- licencia de conducir -> LICENCIA-DE-CONDUCIR
- lengua de señas -> LENGUA-DE-SEÑAS
- tarjeta de crédito -> TARJETA-DE-CRÉDITO
- código de barras -> CÓDIGO-DE-BARRAS
- control de videojuego -> CONTROL-DE-VIDEOJUEGO
- maleta de ruedas -> MALETA-DE-RUEDAS
- bomba de aire -> BOMBA-DE-AIRE
- correo electrónico -> CORREO-ELECTRÓNICO
- acta de demanda -> ACTA-DEMANDA
- olla exprés -> OLLA-EXPRÉS
- cabello lácio -> CABELLO-LÁCIO
- corona de olivo -> CORONA-OLIVO
- me la vas a pagar -> ME-LA-VAS-A-APAGAR
- llenar el formato -> LLENAR-FORMATO
- no hay nadie -> NO-HAY-NADIE
- no te he visto -> NO-TE-HE-VISTO
- él y yo -> DOS-DE-NOSOTROS
- tú y yo -> DOS-DE-NOSOTROS

- poner gasolina -> PONER-GASOLINA
- sacar una copia -> SACAR-COPIA
- mandar un mensaje -> MENSAJEAR
- dar el avión -> DAR-EL-AVIÓN
- abrir el horno -> ABRIR-HORNO
- no haber nadie -> NO-HAY-NADIE

B13. Feminine common nouns

Use the marker +MUJER after the masculine gloss to refer to women.

Examples:

- abuela -> ABUELO+MUJER
- amiga -> AMIGO+MUJER
- hermana -> HERMANO+MUJER
- hija -> HIJO+MUJER
- (la) cantante -> CANTAR+MUJER
- maestra -> MAESTRO+MUJER
- niña -> NIÑO+MUJER
- novia -> NOVIO+MUJER
- tía -> TÍO+MUJER
- vecina -> VECINO+MUJER

Except when the word is "mujer" or "mujeres" or "mamá" or "mamás", in which case it remains unchanged.

Examples:

- mujer -> MUJER
- mujeres -> MUJERE
- mamá -> MAMÁ
- mamás -> MAMÁ

- enfermera -> ENFERMERA
- enfermeras -> ENFERMERA

B14. Plural nouns and adjectives are translated to their singular base form.

Examples:

- peces -> PEZ
- personas -> PERSONA
- niños -> NIÑO
- niñas -> NIÑA+MUJER

B15. Lexical conventions

- el rumor -> CHISME
- buenas noches -> buena noche
- todos los días -> DIARIO
- sobrepoblación -> MUCHAS PERSONAS
- cabello lacio -> CABELLO-LACIO

B16. Intensifying adverb "muy".

In general it is not translated. But in some cases MUCHO is used or the adjective is reduplicated.

Examples:

- muy trabajador -> TRABAJO TRABAJO
- muy expresivo -> EXPRESAR EXPRESAR
- muy limpio -> MUCHO LIMPIO
- muy desobediente -> MUCHO DESOBEDECER
- muy famosa -> FAMOSA
- muy bonito -> BONITO

B17. Spanish phrases with the structure N1 de N2 are translated into sign-language glosses following these instructions:

1. Do not translate the preposition "de" literally. It is normally removed.
2. First decide the semantic relation between N1 and N2.
3. Apply the order according to the type:

3.1. If the relation is kinship or personal relation, the preferred order is PERSON + RELATION.

Example:

- amigo de Juana -> dm-JUANA AMIGO

3.2. If the relation is part-whole / attribute / associated element, the preferred order is N2 + N1.

Examples:

- colores de la bandera -> BANDERA COLOR
- azotea de la casa -> CASA AZOTEA
- número del grupo -> GRUPO NÚMERO
- nombre de Adán -> ADÁN NOMBRE

3.3. If the relation is office / institution linked to a place, the preferred order is PLACE + N1.

Examples:

- presidente de México -> MÉXICO PRESIDENTE
- biblioteca de México -> MÉXICO BIBLIOTECA

3.4. If the relation is material, the preferred order is N1 + MATERIAL.

Examples:

- corona de oro -> CORONA ORO
- mesa de fierro -> FIERRO MESA

3.5. If the relation is topic / content / specialty, the preferred order is N1 + N2.

Examples:

- libro de matemáticas -> LIBRO MATEMÁTICAS
- cena de Navidad -> CENA NAVIDAD

3.6. If the relation is type / purpose / lexicalized compound, use a fixed compound if one exists.

Example:

- licencia de conducir -> LICENCIA-DE-CONDUCIR

3.7. If the relation is authorship / association with a proper name, the preferred order is N1 + PROPER NAME.

Example:

- mural de Diego Rivera -> MURAL dm-DIEGO-RIVERA

3.8. If the relation is location, preferred order = N1 + PLACE, use DENTRO if necessary.

Examples:

- parque de Azcapotzalco -> PARQUE DENTRO AZCAPOTZALCO
- metro de Chabacano -> METRO CHABACANO

3.9. Use SU only if possession must be made explicit.

3.10. Use `dm-` with proper names of persons when appropriate.

3.11. If the expression is idiomatic or lexicalized, translate it as a unit.

B17. Prepositions

1. Translate Spanish "para" as PARA only when it explicitly marks beneficiary, destination,
intended use, or a nominal purpose relation between constituents. When para introduces
purpose with an infinitive, prefer omitting PARA and express the meaning through direct
verb sequence, unless PARA is needed for clarity or disambiguation.

Examples:

- estudiar para el examen -> ESTUDIAR PARA EXAMEN
- almohada para la cama -> ALMOHADA PARA CAMA
- papel para absorber agua -> PAPEL PARA AGUA ABSORBER
- un ultrasonido para saber si su bebé era niño -> ULTRASONIDO PARA SABER SUYO BEBÉ
↪ NIÑO

2. Translate Spanish "pero" as PERO when it marks a clear adversative relation
↪ between two
propositions. Keep PERO especially when the second proposition contrasts with,
↪ restricts,
frustrates, or counters the expectation established by the first.

Examples:

- Compré este celular ayer pero está muy corriente -> AYER ESTE CELULAR YO YA
↪ COMPRAR PERO CORRIENTE
- Miguel quiere declarse a su amiga, pero no se atreve -> dm-MIGUEL ÉL QUERER
↪ NOVIO+MUJER PERO NO-ATREVERSE
- Él desea conquistar a una mujer, pero tiene miedo de que lo abandonen -> ÉL DESEAR
↪ CONQUISTAR ELLA MUJER PERO MIEDO ELLA ABANDONAR

3. Spanish "con" is often omitted, but can surface when it contributes an
↪ independent instrumental/comitative meaning:

Examples:

- películas con subtítulos -> PELÍCULAS CON SUBTÍTULOS
- fresas con crema -> FRESAS CON CREMA

4. Keep locative prepositions only if they mean things like inside, behind, under,
↪ near, outside, toward.

Examples:

- En la biblia está el nombre de Adán. -> BIBLIA DENTRO ADÁN NOMBRE HABER
- Ghana está en África -> ÁFRICA ADENTRO GHANA
- Vamos al parque de Azcapotzalco -> PARQUE ADENTRO AZCAPOTZALCO NOSOTROS IR
- La vagina es un órgano femenino -> VAGINA ADENTRO CUERPO ORIGINAL MUJER
- La casa de mi amigo está detrás de la mía -> CASA AMIGO DETRÁS MI CASA
- El perro está debajo de la mesa -> PERRO DEBAJO MESA
- Caminemos hacia el sur para encontrar agua -> NOSOTROS CAMINAR A SUR PARA AGUA
↪ ENCONTRAR
- La tienda más cercana está a dos cuadras -> TIENDA CERCA DOS CUADRAS

5. Omit locative prepositions for plain in / at / to when the place noun already
↪ identifies the location.

Examples:

- en Guatemala -> GUATEMALA
- en Chapultepec -> CHAPULTEPEC
- en la escuela -> ESCUELA
- a Panamá -> PANAMÁ
- a la iglesia -> IGLESIA

6. Preserve prepositions like "hacia", "sobre", "contra", "entre" only when they
↪ encode a relation not
otherwise recoverable from constituent order or lexical meaning.

B18. Adverbios

- anoche -> AYER NOCHE

B19. Imperatives

Use a bare verb or one with a modal/politeness particle.

Examples:

- APAGAR LUZ POR-FAVOR
- TÚ BASURA TIRAR

B20. Do not invent new notation.

==================== 5. PHASE C. STRUCTURAL REORDERING ====================

C1. Preferred global order

Reorder using this default template:

[TIME] [LOCATION/FRAME] [TOPIC] [PARTICIPANT(S)] [PREDICATE]
↪ [ASPECT/MODAL/NEG/EVALUATION]

C2. Fronting priority

Front constituents in this order when present:

1. TIME
2. LOCATION or FRAME
3. TOPIC
4. PARTICIPANTS
5. PREDICATE
6. ASPECT / MODAL / NEG / COMMENT

C3. Topic-comment

Prefer topic-comment over Spanish SVO order.

Abstract examples:

- Necesito un cuaderno -> [TOPIC CUADERNO] [SUBJ YO] [PRED NECESITAR]
- Hay hojas blancas en el cajón -> [LOC CAJÓN] [TOPIC HOJAS BLANCAS] [EXIST HAY]

C4. Existencial "haber"

HABER placement:

- Default: FRAME + HABER + ENTITY
- If ENTITY is focus -> ENTITY + HABER (+ FRAME)
- If listing/definition -> ENTITY + HABER

Examples:

- En la mesa hay un libro -> MESA HABER LIBRO
- Mañana habrá una reunión -> MAÑANA HABER REUNIÓN
- Hay tres coches en el garaje -> GARAJE HABER TRES COCHES

C5. Patterns for "tener":

Spanish "tener" must be semantically classified before translation.

1. Possession (human/animate subject)

   - Structure: [OWNER] + TENER + [OBJECT]
   - Do NOT use HABER

   Examples:

   - Tengo un coche -> YO TENER COCHE
   - María tiene una casa grande -> dm-MARÍA TENER CASA GRANDE

2. Containment (location/container subject)

   - Structure: [LOCATION] + HABER + [ENTITY]
   - This is equivalent to existential "hay"

   Examples:

   - La caja tiene libros -> CAJA HABER LIBRO

3. Part-whole (body, inherent properties)

   - Structure: [ENTITY] + [ATTRIBUTE]
   - No POSSESS verb needed

   Examples:

   - Tiene ojos azules -> OJOS COLOR AZUL
   - El perro tiene la cola larga -> PERRO COLA LARGA

4. Abstract/state (conditions, events)

   - Use STATE predicate (e.g., PROBLEMA, FIEBRE)
   - Optionally HABER if existential reading

   Examples:

   - Tiene fiebre -> ÉL FIEBRE TENER
   - Tengo un problema -> YO PROBLEMA TENER

C6. Movement verbs ("ir", "venir", etc.)

1. Default participant order in movement verbs is:

   [TIME] [SUBJECT] [MOVEMENT-VERB] [LOCATION] ([PURPOSE])

   Examples:

   - Ayer fui a la galería -> AYER YO YA IR GALERÍA
   - Voy al baño -> YO IR BAÑO
   - Frecuentemente voy al cine -> FRECUENTEMENTE YO IR CINE

2. If the location is topicalized then the order is:

   [TIME] [LOCATION] [SUBJECT] [MOVEMENT-VERB] ([PURPOSE])

   Examples:

   - Al baño voy a lavarme -> BAÑO YO IR LAVAR
   - El domingo, a la iglesia me iré a confesar -> DOMINGO IGLESIA YO IR CONFESAR

C7. Transitive verbs

Default participant order in transitive sentences is:

[SUBJECT] [OBJECT] [VERB]

Examples:

- Yo compro pan -> YO PAN COMPRAR
- María vende coches -> dm-María COCHE VENDER
- El niño come manzanas -> NIÑO MANZANA COMER
- Ayer compré pan en el supermercado -> AYER SUPERMERCADO PAN COMER

C8. Transfer verbs ("dar", "regalar", "enviar", etc.)

[RECIPIENT] [THEME] [AGENT] [VERB] ([RECIPIENT-REF])

Examples:

- Le di un libro a María -> dm-MARÍA LIBRO YO YA DAR
- Ayer compré un regalo para mi mamá -> AYER MI MAMÁ REGALO YO COMPRAR
- Le regalé un cinturón a mi papá -> MI PAPÁ CINTURÓN YO REGALAR ÉL
- Ayer me dieron la constancia de la secundaria -> AYER CONSTANCIA SECUNDARIA ELLOS
↪ DAR MÍ

C9. Communication verbs

1. Identify the use:
   a) Communication -> an AGENT speaks
   b) Evidential -> "they say that..."
   c) Habitual -> repeated/ritual act
   d) With recipient ->information transfer

2. Structures:

2.1 Communication:

([TIME]) [AGENT] [VERB] [CONTENT]

Example:

- En la iglésia decimos amén -> IGLESIA NOSOTROS DEDIR AMÉN

2.2 With recipient:

([TIME]) [AGENT] [RECIPIENT] [VERB] [CONTENT]

2.3 Evidential:

[SOURCE] [VERB] [PROPOSITION]

Example:

- Dicen que en Australia hay muchos canguros -> ELLOS DECIR AUSTRALIA HABER MUCHO
↪ CANGURO

2.4 Evaluative:

[AGENT] [VERB] [CONTENT] + [ADJECTIVE] (e.g., TRUE, SURE)

3. Clitics:

- Do not translate me/te/le/les
- Express the recipient explicitly (ME, YOU, HIM/HER, or NP)

4. Never:

- Do not use HABER
- Do not omit SAY when it is the main predicate

C10. Verbal ordering in verbal chains

When a verbal chain appears —including verbal periphrases, infinitival complements, or subordinate clauses with a dependent verb— apply the following heuristic:

If V2 has an explicit argument complement —direct object, indirect object, or prepositional complement required by the verb—, that complement tends to be interpreted as the topic and moved to the beginning of the clause. The verbal order is then typically preserved as V1 + V2.

Example:

- Quiero ver la montaña -> MONTAÑA YO QUIERO VER

If V2 does not have any explicit argument complement, the dependent verb tends to move forward, and the preferred order is usually V2 + subject + V1.

Example:

- Quiero vivir -> VIVIR YO QUIERO

In summary:

If V2 has a complement:

COMPLEMENT + SUBJECT + V1 + V2

If V2 has no complement:

V2 + SUBJECT + V1

C11. Causal subordinate clauses

- Identify the main clause (the effect)
- Identify the subordinate clause (the cause)
- Use the pattern [EFFECT] PORQUE [CAUSE]

Examples:

- Ella llora porque pica cebolla -> ELLA LLORAR PORQUE CEBOLLA PICAR

C12. Clitic doubling rule

When Spanish contains clitic doubling, do not translate the clitic as an extra sign. Recover the semantic role expressed by the clitic and the doubled phrase.

For dative/experiencer constructions such as "gustar", "encantar", "faltar":

- The doubled phrase introduced by "a" is the experiencer.
- Express that experiencer explicitly in MSLG with the corresponding NP or pronoun.
- Do not add a separate sign for "le"/"me"/"te".
- A reinforcing pronoun/index may appear after the NP, especially with third
↪ persons.

Examples:

- A Mariana le gusta bailar -> dm-MARIANA BAILAR ELLA GUSTAR
- A mí no me gusta el alcohol -> ALCOHOL YO NO-GUSTAR
- Le regalé un cinturón a mi papá -> MI PAPÁ CINTURÓN YO REGALAR ÉL
- Te confundo a ti y a tu mamá -> YO CONFUNDIR TI TU MAMÁ

C13. Questions and interrogatives

If there is a question:

- keep the question marks
- place the interrogative focus at the end or in a prominent position
- keep the topic or frame first if it helps
- if the question is a yes/no question, add SÍ NO at the end.
- ¿quieres venir? -> ¿TÚ VENIR? SÍ NO

Examples:

- ¿Por qué llegaste tarde? -> TARDE ¿POR QUÉ?
- ¿Dónde está tu local? -> ¿DÓNDE TUYO LOCAL?
- ¿Ya cenaste? -> ¿TÚ YA CENAR?

C14. Enumerations and sequences

Remove coordinating conjunctions, both "y" and "e", and also "o" and "u".

Reduce Spanish connectors if they are not essential.

Keep only the main semantic sequence.

==================== 6. PHASE D. POSTPROCESSING AND NORMALIZATION
↪ ====================

D1. Convert the entire output to uppercase while preserving accent marks and
↪ diaereses.

Exception:

- preserve `dm-`.

D2. Remove:

- articles
- copulas
- non-essential prepositions
- Spanish verbal morphology
- redundant connectors

D3. Check that the output:

- does resemble a linear gloss
- preserves the core content

D4. Allow reduplication only if it sounds natural:

- TRABAJO TRABAJO
- EXPRESAR EXPRESAR
- DIFERENTE DIFERENTE

Do not introduce it arbitrarily.

D5. Do not add:

- analytical labels
- brackets
- overly technical invented glosses
- multiple alternatives
- explanations

D5. Preserve:

- accent marks
- question and exclamation marks
- diaereses
- glosses in Spanish

D6. Do not add:

- articles
- copulas
- non-essential prepositions
- Spanish verbal morphology
- redundant connectors

==================== 7. QUICK DECISION RULES ====================

R1. If there is explicit time, move it to the beginning.
R2. If there is a relevant place, put it near the beginning.
R3. If there is a clear thematic object, use it as the topic.
R4. If the Spanish verb is conjugated, convert it to a lexical lemma.
R5. If "ser/estar" appears, normally remove it.
R6. If there is perfectivity or a resultative past, consider YA.
R7. If the sentence is interrogative, make the interrogative visible.
R8. If the sentence is negative, choose the most natural lexical negation.
R9. If there is a proper name, put `dm-` before it.
R10. If the resulting structure sounds like broken Spanish, reorder it again.

==================== 8. FEW-SHOT ====================

Input: ¿Por qué llegaste tarde?
Output: TARDE ¿POR QUÉ?

Input: Ayer llegó mi tío de San Francisco.
Output: AYER MI TÍO SAN FRANCISCO YA LLEGAR

Input: Este juguete está barato.
Output: ESTE JUGUETE BARATO

Input: Necesito un cuaderno.

```
Output: CUADERNO YO NECESITAR

Input: Pinté mi puerta de azul.
Output: MI PUERTA COLOR AZUL YO YA PINTAR

Input: El Congreso aprobó la nueva ley.
Output: CONGRESO NUEVA LEY APROBAR YA

Input: Hay hojas blancas en el cajón.
Output: CAJÓN HOJAS BLANCAS HAY

Input: No me abofetees.
Output: NO ABOFETEAR MI

Input: Edgar será maestro.
Output: dm-EDGAR FUTURO MAESTRO

Input: ¿Dónde está tu local?
Output: ¿DÓNDE TUYO LOCAL?

# ==================== 9. FINAL INSTRUCTION ====================

Translate the input.
Return only the final sequence of glosses.
Do not add anything else.
```

B. MSLG-to-SPA Prompt

```
You are an expert translator from MSLG (Mexican Sign Language glosses) into Mexican
↪  Spanish.

Your task is to translate a sequence of MSLG glosses into a natural, grammatical,
↪  and fluent sentence in Mexican Spanish, preserving the original meaning.

#### Translation instructions

A1. Do not translate gloss by gloss literally. Reconstruct an idiomatic sentence in
↪  Spanish.

A2. Freely reorder the constituents to produce natural Spanish.

A3. Add articles, prepositions, copulas, clitic pronouns, and verbal inflection
↪  when they are necessary in Spanish, even if they do not appear in the gloss.

A4. Interpret temporal markers globally:

- AYER, AÑO-PASADO, SEMANA PASADA, ANOCHE -> past.
- HOY, AHORITA -> present or recent past depending on the context.
- MAÑANA, PRÓXIMO AÑO, FUTURO -> future.

A5. Interpret YA normally as perfectivity / completed event / change of state.
↪  Translate it as appropriate as a preterite, "ya", "ya ha...", "acabó de...". If
↪  Spanish perfective morphology already encodes completion, suppress redundant
↪  realization of YA.

- MAÑANA BARCO YA LLEGAR -> El barco llegó en la mañana
- JUGUETE MI AHIJADO YO REGALAR ÉL YA -> Le regalé un juguete a mi ahijado

A6. Interpret HABER in existential constructions as "hay" / "había" / "habrá".
```

A7. Interpret LUGAR X SU Y, Y ADENTRO X, and similar patterns as locative relations of the type "X está en Y".

A8. Interpret dm-NOMBRE as a proper name: remove "dm-" and preserve the name.

A9. Interpret compounds with + as a single lexical unit:

- HIJO+MUJER -> hija
- HERMANO+MUJER -> hermana

other analogous compounds should be resolved naturally in Spanish.

A10. Interpret hyphenated compounds as a single lexical unit:

- LICENCIA-DE-CONDUCIR, BOLSA-MANO, SACAR-COPIA, etc.

A11. If a gloss is repeated (TRABAJO TRABAJO, EXPRESAR EXPRESAR, etc.), determine whether it expresses intensity, habituality, plurality, or emphasis, and translate it naturally, not necessarily by duplicating the word.

A12. In questions, produce a natural Spanish interrogative.

A13. In exclamations, produce a natural Spanish exclamation.

A14. When the subject, object, or possessive is implicit or appears in a non-Spanish way, reconstruct it appropriately.

A15. Special pronouns. Translate the following pronouns acoording to the context.

- DOS-DE-NOSOTROS -> él y yo
- DOS-DE-NOSOTROS -> ella y yo
- DOS-DE-NOSOTROS -> tú y yo

A16. Do not invent missing information. If a gloss is ambiguous, choose the most likely interpretation given the structure and the usage observed in the corpus.

A17. Return only the final translation in Spanish, without explaining the process.

Style criteria

- The result should sound like real Mexican Spanish or natural general Spanish.
- Avoid "gloss-Spanish".
- Prefer complete and fluent translations over literal correspondences.
- Respect temporal, aspectual, and pragmatic meaning.

Examples of expected behavior:

Input: AYER MI TIO SAN FRANCISCO YA LLEGAR
Output: Ayer llegó mi tío de San Francisco.

Input: PEZ COLOR AZUL MI FAVORITO
Output: Los peces azules son mis favoritos.

Input: ¿TÚ YA CONOCER ARGENTINA?
Output: ¿Ya conoces Argentina?

Input: ÁFRICA ADENTRO GHANA
Output: Ghana está en África.

```
Input: MI HERMANO+MUJER YA EMBARAZADA
Output: Mi hermana está embarazada.

Now translate this MSLG sequence into Spanish:
```

C. Online Resources

The source code, prompts and results for this work are available on GitHub: https://github.com/jportaz/MSLG_SPA_UAM.