

Sistema de Traducción Automática del Español Mexicano a Glosa de la Lengua de Señas Mexicana mediante Aprendizaje Profundo



Colaboración

Cristian M. Torres-Martínez; Vanessa Abigail Huerta-García; Joel A. Trejo-Sánchez, Tecnológico Nacional de México / Ciudad Guzmán; Jesús Ezequiel Molinar-Solis, Centro de Investigación en Matemáticas; Daniel Fajardo-Delgado, Tecnológico Nacional de México / Ciudad Guzmán

Fecha de recepción: 21 de noviembre de 2025

Fecha de aceptación: 21 de noviembre de 2025

RESUMEN: Este trabajo presenta un sistema de traducción automática de texto en español a la glosa de la lengua de señas mexicana (LSM), una representación intermedia y gramatical que refleja la secuencia de señas de esta lengua. Para el desarrollo del sistema, se construyó un corpus bilingüe LSM-español a partir del diccionario de la LSM de la Ciudad de México, complementado con información lingüística adicional y recursos audiovisuales, y estructurado en una base de datos relacional. Posteriormente, utilizando el modelo NLLB de Meta, se realizó un ajuste fino con dicho corpus, generando tres variantes del modelo: ES-LSM1, ES-LSM2 y ES-LSM3. La calidad de cada versión se evaluó con las métricas BLEU, ROUGE y ChrF, observándose mejoras significativas en cada versión, incluso con una cantidad limitada de datos. En particular, la versión ES-LSM3 alcanzó resultados destacados con 84% en ROUGE, 75% en ChrF y hasta 65% en BLEU-1, reflejando traducciones con alta precisión léxica y correspondencia con las frases de referencia. Finalmente, el modelo ES-LSM3 se integró en una interfaz gráfica que permite traducir palabras del español a glosa de la LSM y reproducir videos que muestran brevemente el uso de una seña para cada palabra o, en su defecto, el deletreo dactilológico.

PALABRAS CLAVE: Aprendizaje profundo, compilación de corpus, glosa de la LSM, lengua de señas mexicana, sistema de traducción automática.

ABSTRACT: This work presents a machine translation system for converting Spanish text into glosses of Mexican Sign Language (LSM). The LSM gloss serves as an intermediate grammatical representation that reflects the sequence of signs in this language. For the development of the system, a Spanish-LSM bilingual corpus was built using the LSM dictionary of Mexico City, complemented with additional linguistic information and audiovisual resources, and structured in a relational database. Subsequently, using Meta's NLLB model, fine-tuning was performed with this corpus, resulting in three model variants: ES-LSM1, ES-LSM2, and ES-LSM3. The quality of each version was evaluated using the BLEU, ROUGE, and ChrF metrics, showing significant improvements in each iteration, even with a limited amount of data. In particular, the ES-LSM3 version achieved outstanding results with 84% in ROUGE, 75% in ChrF, and up to 65% in BLEU-1, reflecting translations with high lexical accuracy and strong correspondence to the reference sentences. Finally, ES-LSM3 was integrated into a graphical interface that allows the translation of Spanish words into LSM glosses and playing videos that briefly demonstrate the use of a sign for each word or, when unavailable, its fingerspelled form.

KEYWORDS: Deep learning, corpus compilation, LSM gloss, mexican sign language, automatic translation system.

INTRODUCCIÓN

En México, aproximadamente 2.9 millones de personas presentan algún tipo de discapacidad auditiva, de las cuales cerca de 1.3 millones son sordas [1]. Esta cifra subraya la importancia del conocimiento, uso y difusión de la lengua de señas mexicana (LSM), un sistema de comunicación visual con estructura lingüística propia, basado en configuraciones manuales, movimientos,

expresiones faciales y corporales. A pesar de su papel fundamental en la inclusión social de las personas sordas, únicamente el 0.2% de la población mexicana utiliza o comprende la LSM [2], lo que revela una marcada brecha comunicativa entre personas sordas y oyentes.

La LSM fue reconocida como lengua nacional oficial en 2005 y constituye un elemento central del patrimonio lingüístico de la comunidad sorda [3]. Como toda lengua natural, posee una gramática propia, sustentada en una fonología visual que integra tres componentes: la matriz segmental (movimiento y uso de las manos), la articulatoria (configuración, ubicación, dirección y orientación) y los rasgos no manuales (expresiones faciales y posturas) [4, 5].

Un medio para comprender la expresión del significado en LSM es el uso de la glosa, una transcripción de las señas a la sintaxis del español. Aunque la glosa ofrece una representación aproximada del mensaje original, es una herramienta útil en la enseñanza, el análisis lingüístico y la documentación de la LSM, facilitando su estudio y comparación con otras lenguas de señas [5].

En años recientes, se han desarrollado diversos sistemas automáticos para facilitar la comunicación entre personas sordas y oyentes, destacando aquellos basados en visión computacional y aprendizaje automático para el reconocimiento de señas [6-8]. Sin embargo, la mayoría de estos esfuerzos se enfocan en la traducción de la LSM al español, mientras que los trabajos dirigidos a la traducción del español a la LSM han sido escasos.

Entre estos pocos trabajos, destacan los de Pichardo-Laguna et al. [9,10,11], quienes desarrollaron un sistema de traducción directa con base a reglas marcadas. Un trabajo más reciente de ellos considera la sintaxis tanto del español como del LSM [12]. Por otra parte, Hernández-Cruz [13] propone un traductor del español mexicano a la glosa de la LSM, cuyo diseño combina algunos modelos tradicionales, como autómatas y reglas lingüísticas, con una red de memoria a largo plazo (LSTM). Sin embargo, el corpus empleado para el entrenamiento proviene principalmente de traducciones del español a la glosa de la lengua de señas española, lo que introduce un sesgo que podría afectar su aplicabilidad a la LSM. En los últimos años, han surgido enfoques basados en aprendizaje profundo que podrían superar estas limitaciones al aprovechar corpus más extensos y técnicas de modelado del lenguaje, e.g., ver [14]. Un ejemplo reciente es el trabajo de Lara-Ortiz et al. [15], quienes compilan un corpus de 3,000 pares de frases en español y glosa de LSM, validado con modelos transformadores y alcanzando una efectividad del 94%. No obstante, su corpus presenta una redundancia de estructuras y de vocabulario, lo que podría afectar la generalización de los modelos.

El presente trabajo propone un sistema de traducción automática del español mexicano a la glosa de la LSM.

Entre los objetivos de este trabajo destacan tres. En primer lugar, desde el punto de vista científico, se planteó analizar la viabilidad del uso de modelos multilingües de código abierto para abordar tareas de traducción entre lenguas orales y lenguas de señas, considerando los desafíos lingüísticos y estructurales que implica este tipo de procesamiento. En segundo lugar, desde una perspectiva tecnológica, el objetivo fue desarrollar un prototipo funcional basado en modelos de aprendizaje profundo capaz de generar traducciones bidireccionales entre español y glosa de LSM, evaluando su desempeño con un conjunto de validación independiente. Finalmente, desde un enfoque social, se buscó contribuir a la inclusión comunicativa y a la reducción de barreras lingüísticas que enfrenta la comunidad sorda en México, mediante el aprovechamiento de tecnologías accesibles y de libre uso.

En conjunto, el trabajo presenta dos contribuciones principales: (1) el preprocesamiento de un corpus bilingüe LSM-español compuesto por 1,077 pares de frases únicas, cada una con vocabulario distinto, y (2) el desarrollo de tres modelos de traducción ajustados a partir de No Language Left Behind (NLLB) de Meta, un modelo multilingüe de código abierto que ha demostrado un desempeño competitivo en tareas de bajo recurso. Los resultados experimentales indican un desempeño de hasta 69% de precisión en la traducción de glosa de LSM al español, y 68% en la dirección inversa, lo que evidencia la viabilidad técnica y científica del enfoque propuesto, así como su potencial impacto social en el fortalecimiento de la accesibilidad lingüística.

Revisión de la literatura

A la fecha, se han documentado poco más de 50 trabajos en la literatura científica centrados en la traducción automática entre la LSM y el español mexicano. La mayoría de estos estudios abordan el reconocimiento automático de señas, utilizando imágenes y videos en dos o tres dimensiones para capturar la configuración, el movimiento, la orientación y la ubicación de las manos [16-18]. Algunos enfoques han incorporado sensores inerciales para mejorar la precisión en la detección de señas, como en el caso de Varela-Santos et al. [19]. Particularmente, Ríos-Figueroa et al. [16] presentan un compendio de trabajos destacados en esta área, mientras que Mendoza-Tene et al. [6] evalúan la efectividad de estos sistemas en dispositivos móviles.

Por otra parte, la traducción automática en la dirección inversa, del español mexicano a la LSM, ha recibido menor atención. Entre los estudios más relevantes se encuentran los de Pichardo-Lagunas y Martínez-Seis [9,10], quienes propusieron un enfoque basado en reglas lingüísticas manuales para estructurar la LSM a partir del español. Estos autores identificaron que muchas construcciones requieren referencias visuales contextuales, como en el caso de las preposiciones de lugar, para lograr una equivalencia semántica adecuada. Así mismo,

observaron que, al igual que en otras lenguas, la LSM permite múltiples estructuras sintácticas para expresar una misma idea. Por su parte, Hernández-Cruz [13] desarrolló un sistema de traducción del español mexicano a la glosa de la LSM, combinando autómatas, reglas gramaticales y una red neuronal LSTM. Sin embargo, su corpus de entrenamiento se basó principalmente en traducciones a la lengua de señas española, lo que introduce un sesgo significativo para su aplicabilidad en el contexto mexicano.

En contraste, otras lenguas de señas han mostrado avances importantes. Por ejemplo, Zhao et al. [20] desarrollaron un sistema de traducción de texto a la lengua de señas americana (ASL) utilizando gramáticas de árbol sincrónico. Del mismo modo, D'Armond y Speers [21] emplearon gramáticas léxicas funcionales para representar estructuras del inglés en ASL, integrando además un modelo fonológico basado en el principio de movimiento-retención. También se han propuesto sistemas para otras lenguas de señas, como la española [22], la griega [23] y la árabe [24], entre otras [14].

En años recientes, los avances en aprendizaje profundo, particularmente los modelos de lenguaje de gran escala, han transformado los sistemas de traducción automática [25,26]. Modelos como BERT [27] y sus variantes, entre ellas XLNet [28] y RoBERTa [29], han establecido nuevos estándares en tareas de procesamiento del lenguaje natural. En este sentido, Cate y Hussein [30] exploraron el uso de modelos tipo BERT para la traducción del inglés a ASL con resultados prometedores. De forma complementaria, Stoll et al. [31] propusieron redes generativas adversarias (GAN) para mejorar la generación de señas. Sin embargo, un obstáculo persistente para el desarrollo generalizado de sistemas de traducción automática hacia lenguas de señas sigue siendo la escasez de corpus extensos, balanceados y bien estructurados [32].

MATERIAL Y MÉTODOS

Construcción del Corpus

El corpus utilizado en este estudio se construyó principalmente a partir del Diccionario de la Lengua de Señas Mexicana de la Ciudad de México [4], el cual contiene 977 pares de sentencias con su transcripción en glosa de la LSM y su correspondiente en español. Cada sentencia incorpora una palabra nueva del diccionario, asegurando así una cobertura progresiva del vocabulario. Adicionalmente, se integraron 100 palabras individuales provenientes de otras fuentes, con el objetivo de ampliar la diversidad léxica. En total, el corpus está compuesto por 1077 registros.

Además de las transcripciones LSM-español, se incluyeron anotaciones lingüísticas detalladas de la LSM, abarcando los siguientes aspectos:

- Tipo de seña: Se clasifican en seña manual (SM), cuando se emplea únicamente la mano dominante; seña bimanual (SB), cuando participan la mano dominante y la mano base; y seña simétrica (SS), cuando ambas manos se

usan de forma sincronizada. También se consideran las señas compuestas (SC), integradas por dos o más señas simples que, en conjunto, representan un único concepto. Estas pueden incluir combinaciones de SM, SB o SS.

- Configuración: Hace referencia a la forma que adoptan las manos al realizar la seña. Se seleccionaron configuraciones previamente registradas en el diccionario, así como algunas configuraciones transitivas, es decir, aquellas que cambian de forma durante la ejecución de una misma seña.

- Orientación: Indica la dirección en la que se orienta la palma o el dorso de la mano. Algunas señas presentan orientaciones fijas, mientras que otras implican un cambio de orientación como parte de su significado.

- Ubicación: Corresponde al espacio corporal donde se inicia, desarrolla o finaliza la seña. Por ejemplo, una seña puede comenzar a la altura del pecho y terminar en la frente, o viceversa. La ubicación varía ampliamente entre señas.

- Movimiento: Describe el desplazamiento que realiza una o ambas manos. En algunos casos, una mano permanece estática mientras la otra realiza un movimiento; en otros, ambas manos interactúan dinámicamente.

- Rasgos no manuales: Se documentan elementos como expresiones faciales, movimientos de cabeza, cejas u hombros que complementan o modifican el significado de la seña.

- Representación icónica: En los casos en que la seña simula visualmente una acción, objeto o ser vivo, se incluye una nota de representación para señalar dicha correspondencia.

- Definición: Se anota la definición en español de cada palabra registrada, siguiendo la referencia del diccionario. Con el fin de complementar el corpus textual, se procuró incorporar material audiovisual de las señas. Aunque no todas las entradas disponen de un video demostrativo, se logró recopilar al menos una representación en video para la mayoría de los registros.

Finalmente, el corpus fue estructurado en una base de dato relacional implementada en MySQL, lo que permite un almacenamiento organizado y eficiente de la información lingüística y visual. La Figura 1 muestra el diagrama entidad-relación de dicha base de datos.

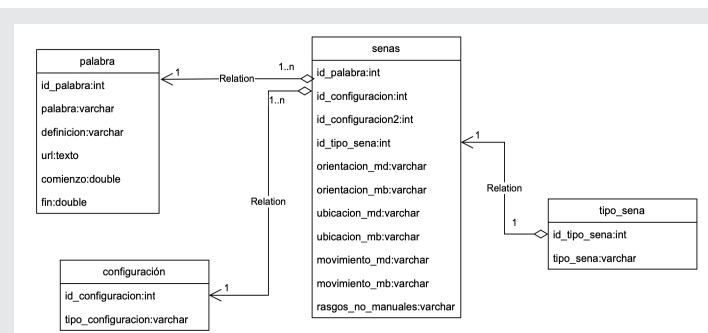


Figura 1: Método propuesto para la clasificación de enfermedades de las hojas de fresa.

Fuente: Elaboración propia.

Entrenamiento de los modelos

Para el entrenamiento del sistema de traducción glosa LSM-español se utilizó el modelo No Language Left Behind (NLLB) desarrollado por Meta. Este modelo se seleccionó sobre otras alternativas (como MarianMT o mBART) por dos razones principales: (1) NLLB está diseñado para abordar la traducción multilingüe de más de 200 idiomas y se ha destacado por una mayor calidad de traducción, en comparación con modelos de arquitecturas codificador-decodificador; y (2) NLLB se destaca por su eficiencia y velocidad en procesos de ajuste fino (fine-tuning), lo que permite adaptar el modelo pre-entrenado con un consumo de recursos computacionales moderado. En particular, en este estudio se empleó una versión optimizada y de menor tamaño (o destilada) del modelo, con un nuevo par de lenguajes: español y glosa de la LSM.

El proceso de entrenamiento se realizó en Google Colab, tanto en su versión gratuita como de paga y constó de tres etapas principales:

1. Preparación del corpus. El corpus se organizó en un archivo CSV con tres columnas: sentencias en español, su correspondiente traducción en glosa LSM, y una indicación del uso de la instancia (como entrenamiento o prueba). Se utilizó el modelo conjunto-validación, considerando 90% para entrenamiento y el resto para pruebas o validación. Se consideró como criterio de división del conjunto-validación la estratificación, asegurando una representación proporcional de la complejidad lingüística en ambos subconjuntos. La estratificación se diseñó bajo dos escenarios con el fin de evaluar la robustez del modelo: (1) un escenario en el que el conjunto de validación compartiera la mayor parte del vocabulario presente en el entrenamiento, permitiendo medir la capacidad de generalización sobre patrones ya observados; y (2) un escenario en el que el conjunto de validación incluyera glosas y estructuras ausentes en el entrenamiento, con el propósito de evaluar la respuesta del modelo ante patrones novedosos. A partir de esta preparación se generaron las distintas versiones del sistema de traducción.

2. Configuración del entorno de entrenamiento. Se añadió un token específico para la glosa LSM al tokenizador del modelo. Este token se inicializó con pesos derivados del español para mejorar la adaptación, evitando así el entrenamiento desde cero y acelerando la convergencia del modelo. Además, se establecieron los hiperparámetros adecuados para el proceso de entrenamiento, en función de la capacidad computacional disponible y el tamaño del corpus. Esta configuración incluyó el tamaño de lote (batch size), número de pasos de calentamiento (warmup steps), longitud máxima de las secuencias y tasa de decaimiento de los pesos (weight decay).

3. Entrenamiento del modelo. El modelo se entrenó utilizando las instancias del corpus. La duración del entrenamiento varió según los recursos asignados (como la capacidad en GPUs) y los hiperparámetros utilizados. El entrenamiento se llevó a cabo en ciclos ajustados a las limitaciones del entorno.

Se construyeron tres modelos de traducción automática, todos ellos versiones del modelo NLLB con diferentes configuraciones. El primer modelo, llamado ES-LSM1, fue entrenado en la versión gratuita de Google Colab, utilizando una GPU T4 de 15 GB. Debido a las limitaciones de hardware, el batch size se restringió a 16. El modelo se entrenó con 927 instancias y alcanzó 19,000 pasos antes de ser interrumpido por el límite de tiempo de cuatro horas impuesto por la plataforma.

El segundo modelo, ES-LSM2, fue entrenado en Google Colab Pro, con acceso a una GPU A100 de 80 GB. Esto permitió aumentar el batch size a 32 y ejecutar 57,000 pasos de entrenamiento, utilizando 977 instancias. El entrenamiento duró aproximadamente seis horas y arrojó mejores resultados en las métricas de evaluación.

Finalmente, para el tercer modelo, ES-LSM3, se amplió el corpus a 1077 instancias, incorporando palabras individuales además de sentencias, con el objetivo de analizar el comportamiento del modelo ante vocabulario previamente aprendido. Se mantuvieron las configuraciones del segundo modelo, permitiendo una comparación directa.

La Tabla 1, muestra los hiperparámetros utilizados en el entrenamiento para los modelos ES-LSM1, ES-LSM2 y ES-LSM3. Para todos los modelos, se estableció una tasa de decaimiento de los pesos del 0.001 y un número de pasos de calentamiento igual a 1000.

Tabla 1. Hiperparámetros para el entrenamiento de los modelos ES-LSM1, ES-LSM2 y ES-LSM3.

	ES-LSM1	ES-LSM2	ES-LSM3
Tamaño de lote	16	32	32
Longitud máxima de las secuencias	128	128	128
Pasos de entrenamiento	19000	57000	57000
Instancias de entrenamiento	927	977	1077

Fuente: Elaboración propia.

La Figura 2 presenta un diagrama de flujo metodológico que sintetiza las fases secuenciales implementadas para el entrenamiento del sistema de traducción glosa LSM-español.

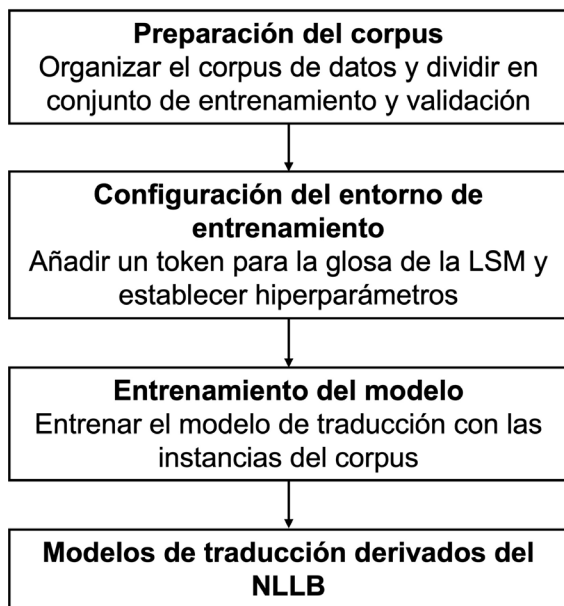


Figura 2: Diagrama de flujo metodológico del proceso de entrenamiento de los sistemas de traducción basados en NLLB. Fuente: Elaboración propia.

Interfaz de lectura de la glosa de la LSM

Se desarrolló una interfaz que, utilizando los modelos de traducción automática presentados en la sección de “Entrenamiento de los modelos”, permite recibir un texto en español mexicano y traducirlo a una secuencia de videos correspondientes a la LSM (ver Figura 3). Estos videos representan de forma directa la glosa LSM generada por el modelo de traducción.



Figura 3. Interfaz del sistema de traducción para convertir texto del español mexicano a videos correspondientes en LSM. Fuente: Elaboración propia.

El software implementado consistió en una aplicación de escritorio desarrollada en Python, la cual opera mediante la búsqueda de cada palabra de la oración ingresada dentro de la base de datos de la Figura 1. Si la palabra está registrada, el sistema reproduce el video correspondiente a su seña. En caso de que la palabra no se encuentre en el repositorio, el sistema recurre al deletreo dactilológico, asegurando así la continuidad del mensaje y facilitando la comunicación con personas sordas.

RESULTADOS

Se realizó un estudio experimental con el propósito de comparar el desempeño de los modelos de traducción generados. Una vez completado el proceso de entrenamiento, se evaluó la capacidad de los modelos para traducir en ambos sentidos: del español a la glosa de la LSM y viceversa. La evaluación se llevó a cabo utilizando métricas estándar como BLEU, ROUGE y ChrF, las cuales permiten medir la similitud entre las traducciones generadas y textos de referencia. BLEU se enfoca en coincidencias n-gram para evaluar precisión, ROUGE valora la recuperación de información en resúmenes y traducciones, y ChrF analiza similitud a nivel de caracteres, siendo útil para lenguas morfológicamente complejas. Estas métricas proporcionan una visión integral del rendimiento de los modelos desarrollados.

La Tabla 2, muestra los datos comparativos de los modelos de la traducción del español a la glosa de la LSM considerando todo el vocabulario dentro del corpus. Se observa, en general, una mejora progresiva en el rendimiento de los modelos a través de sus versiones 1, 2 y 3. lo que indica una mejor calidad de traducción en la versión más reciente. En particular, ES-LSM3 obtuvo los puntajes más altos en ROUGE (84%), ChrF (75%) y BLEU-1 a BLEU-4, lo que indica mayor fidelidad, coherencia y precisión en las traducciones.

Tabla 2. Comparación del desempeño de los modelos en la traducción del español a glosa de la LSM, considerando todo el vocabulario disponible en el corpus.

	ES-LSM1	ES-LSM2	ES-LSM3
ROUGE	82%	83%	84%
ChrF	73%	72%	75%
BLEU 1	52%	52%	65%
BLEU 2	38%	36%	50%
BLEU 3	28%	29%	37%
BLEU 4	23%	22%	29%

Fuente: Elaboración propia.

La Tabla 3 muestra los datos comparativos de los modelos de la traducción del español a la glosa de la LSM considerando solamente parte de vocabulario del cor-

pus. Aquí se observa un comportamiento mixto entre los modelos. ES-LSM1 presenta los valores más altos en ROUGE (83%) y ChrF (69%), lo que indica una mejor similitud entre las sentencias. Sin embargo, ES-LSM3 superó a los otros modelos en las métricas BLEU-2 a BLEU-4, lo que sugiere una mayor precisión en secuencias más largas de palabras.

Tabla 3. Comparación del desempeño de los modelos en la traducción del español a glosa de la LSM, considerando únicamente parte del vocabulario disponible en el corpus.

	ES-LSM1	ES-LSM2	ES-LSM3
ROUGE	83%	81%	74%
ChrF	69%	67%	65%
BLEU 1	68%	71%	69%
BLEU 2	48%	45%	50%
BLEU 3	35%	31%	37%
BLEU 4	25%	22%	29%

Fuente: Elaboración propia.

La Tabla 4 muestra los datos comparativos de los modelos de la traducción de la glosa de la LSM al español considerando todo el vocabulario incluido en el corpus. Aquí, los resultados indican una mejora del modelo ES-LSM2 en casi todas las métricas, especialmente en ROUGE (77%), ChrF (65%) y BLEU-1 a BLEU-4. Aunque ES-LSM3 sigue manteniendo un desempeño competitivo, ES-LSM1 muestra el rendimiento más bajo, especialmente en ChrF (39%) y BLEU-1 (32%).

Tabla 4. Comparación del desempeño de los modelos en la traducción de glosa LSM al español, considerando todo el vocabulario disponible en el corpus.

	ES-LSM1	ES-LSM2	ES-LSM3
ROUGE	69%	77%	76%
ChrF	39%	65%	60%
BLEU 1	32%	60%	50%
BLEU 2	29%	50%	45%
BLEU 3	26%	45%	41%
BLEU 4	22%	35%	36%

Fuente: Elaboración propia.

Finalmente, la Tabla 5 presenta los datos comparativos de los modelos de la traducción de la glosa de la LSM al español, considerando solamente parte del vocabulario del corpus. Similar a la Tabla 1, aquí se observa que ES-LSM3 presenta el mejor desempeño general, destacando en todas las métricas evaluadas.

Tabla 5. Comparación del desempeño de los modelos en la traducción de glosa LSM al español, considerando solamente parte del vocabulario disponible en el corpus.

	ES-LSM1	ES-LSM2	ES-LSM3
ROUGE	74%	72%	74%
ChrF	74%	73%	77%
BLEU 1	63%	62%	68%
BLEU 2	54%	52%	61%
BLEU 3	46%	46%	54%
BLEU 4	36%	39%	49%

Fuente: Elaboración propia.

Análisis cuantitativo

Durante la experimentación con las distintas versiones del modelo, se observó una mejora progresiva en el desempeño conforme se aumentaron las instancias de entrenamiento. Las evaluaciones se realizaron en dos escenarios: un conjunto de pruebas con palabras presentes en el corpus original y otro con palabras nuevas no vistas durante el entrenamiento, siendo el porcentaje de palabras en ambos casos prácticamente el mismo. En ambos casos, se utilizaron métricas estándar de traducción automática (BLEU, ROUGE y ChrF) para medir la calidad de las traducciones. Tomando en cuenta los todos los resultados obtenidos por las diferentes métricas, en los dos casos el desempeño promedio del modelo ES-LSM1 es de 49.5%, el del modelo ES-LSM2 de 53.63 % y el modelo ES-LSM3 tiene un promedio de 56.46%.

Los resultados muestran que las versiones sucesivas del modelo (ES-LSM2 y ES-LSM3) superaron consistentemente a la versión inicial (ES-LSM1). En particular, la versión ES-LSM3 destacó en los puntajes BLEU-4 y ChrF, lo cual indica una mayor precisión y fluidez tanto en traducciones conocidas como no vistas. Esto sugiere que el modelo no solo aprende a reproducir las estructuras del corpus, sino que también adquiere capacidad de generalización. Aunque otros trabajos previos a este hayan utilizado las mismas métricas de evaluación de traducción, no sería del todo justo comparar los puntajes como tal, ya que ninguno ha tenido la misma estructura que utilizamos para entrenar y probar los modelos. Pero si se puede afirmar que nuestra glosa es de la más diversa, lo que resultó en un modelo más adaptable a diferentes traducciones.

Al integrar este modelo con la interfaz de usuario, se logró una herramienta funcional que permite traducir sentencias en español a glosa LSM y reproducir automáticamente los videos correspondientes a cada señal. Estos avances reflejan tanto la viabilidad técnica como el potencial social del sistema desarrollado.

Análisis cualitativo

El modelo ES-LSM1 incluyó varias predicciones acertadas y otras menos precisas, aunque cercanas al mensaje original. De este modelo, ciertas predicciones resultaron más intrigantes que otras, por lo que se requirió una revisión particular de aquellas más alejadas del mensaje esperado, para determinar las causas de dichas variaciones.

Por ejemplo, para la oración “Mi familia siempre dona a Teletón”, la traducción esperada es: “SIEMPRE MI FAMILIA TELETÓN DONAR”, pero la predicción final resultó: “SIEMPRE MI FAMILIA DONAR TELETÓN”. Aunque el modelo fue capaz de identificar y traducir correctamente los lexemas (palabras individuales), tendió a reproducir la estructura canónica Sujeto-Verbo-Objeto (SVO) del español, en lugar de adoptar las estructuras propias de la LSM, las cuales a menudo priorizan el tópico o emplean órdenes como Tiempo-Sujeto-Objeto-Verbo (T-S-O-V).

También se detectaron casos en los que la traducción no fue correcta, especialmente en ciertas palabras. Por ejemplo, para la oración “Mi viaje a Francia me emocionó”, la traducción correcta sería: “YO VIAJE FRANCIA YO EMOCIONAR YA”; sin embargo, la predicción del modelo fue: “MI VIAJE FRANCIA YO EMUCIONAR”. Aquí se observa que las palabras ausentes en el conjunto de entrenamiento no son traducidas adecuadamente, debido a la falta de referencias suficientes para el modelo. Por lo tanto, se determinó que el modelo requería un mayor volumen de información para su entrenamiento, lo que motivó la creación de los modelos ES-LSM2 y ES-LSM3.

La Tabla 6 muestra algunas traducciones del modelo ES-LSM2, donde se evidencia una clara mejoría respecto al modelo ES-LSM1, aunque todavía persisten problemas relacionados con la morfología y la concordancia. El modelo luchó por integrar y predecir correctamente los elementos gramaticales que marcan la persona, como los pronombres clíticos y las señales de direccionalidad inherentes a la LSM.

Tabla 6. Ejemplos comparativos de traducciones generadas por el modelo ES-LSM2, mostrando la calidad de salida frente a las referencias del corpus

Palabra en español	Traducción original LSM	Predicción LSM modelo
Te ayudo	AYUDAR	yo-AYUDAR-a
Le dije a Juan	DECIR JUAN	JUAN YO DEC
Felipe es inteligente	FELIPE INTELIGENTE	FELIPE INTELIGENTE
Gordísimo	GORDO MUCHO	GordisIMO
Había dos bicicletas	BICICLETAS DOS	HAY BICICLETAS HABER

Fuente: Elaboración propia.

La Tabla 7 muestra los resultados de traducción del modelo ES-LSM3. Aquí, el modelo mostró una capacidad mejorada para generalizar y manejar vocabulario desconocido, demostrando un progreso en la inferencia léxica (e.g., intentando traducir palabras no vistas, aunque con errores como “MATINAR” por “MATAR”). Al igual que en los otros modelos, se evidencia la necesidad de un mayor volumen de datos para un entrenamiento más robusto y de alta calidad para abordar estas variaciones y mejorar la precisión tanto léxica como sintáctica.

Tabla 7. Ejemplos comparativos de traducciones generadas por el modelo ES-LSM3, mostrando la calidad de salida frente a las referencias del corpus.

Palabra en español	Traducción original LSM	Predicción LSM modelo
Visité CDMX	YO YA VISITAR CDMX	CDMX YO VISITAR
No hay agua	AGUA NO-HAY	AGUA NO-HAY
Pepe no comió	PEPE NO COMER	PEPÉ NO COM
Pedro mató a Juan	PEDRO MATAR JUAN	JUAN PE MATINAR
Mi hijo es maestro	MI HIJO ÉL MAESTRO	MI HIJO MAES

Fuente: Elaboración propia.

CONCLUSIONES

En el presente trabajo, se desarrolla un sistema de traducción automática de texto en español a glosa de la LSM, con reproducción audiovisual de las señas. Para ello, se construye un corpus bilingüe a partir del diccionario de LSM de la Ciudad de México, enriquecido con información lingüística y representación audiovisual, y organizado en una base de datos relacional. Se proponen tres versiones derivadas del modelo NLLB de Meta, que demuestran mejoras progresivas en el desempeño conforme el incremento de las instancias de entrenamiento. Las métricas de evaluación (BLEU, ROUGE y ChrF) reflejan una mayor calidad de traducción en la versión ES-LSM3, que exhibe una mejor capacidad de generalización, incluso con datos no vistos durante el entrenamiento.

En general, los resultados obtenidos son alentadores, especialmente considerando la limitada cantidad de datos disponibles. Si bien los modelos de traducción automática suelen requerir decenas de miles de sentencias para lograr buenos desempeños, los modelos generados permiten una versión funcional de traducción. Se espera que estos avances representen un paso hacia el fortalecimiento de la educación inclusiva y la accesibilidad tecnológica, al facilitar herramientas que contribuyan a reducir barreras comunicativas para la comunidad usuaria de la LSM.

Como trabajo futuro, se priorizará la expansión del corpus con aportaciones de expertos y usuarios. Fundamentalmente, se realizará una validación con personas sordas e intérpretes de LSM para evaluar la comprensión semántica y la naturalidad de las traducciones producidas por el modelo. Asimismo, se plantea como línea alternativa de investigación el uso de técnicas de minería de patrones frecuentes y reglas de asociación para identificar nuevas reglas de traducción que complementen los modelos propuestos y contribuyan a mejorar la calidad de las traducciones.

AGRADECIMIENTOS

La presente investigación fue financiada parcialmente por el Tecnológico Nacional de México (TecNM) a través del proyecto 23018.25-P, bajo la dirección del Dr. Fajardo-Delgado. Asimismo, el Dr. Molinar-Solis agradece al TecNM por su apoyo para la realización de su Año Sabático bajo el Programa de Investigación Científica, Tecnológica o Educativa (No. de autorización AS-1-043/2025). Además, los autores agradecen el apoyo técnico de Abril Carolina Díaz Magaña.

BIBLIOGRAFÍA

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

[2] Cámara de Diputados (2021). *Aprueban reformas para que personas con discapacidad auditiva reciban educación bilingüe en lengua de señas*. Boletín No.5854 de la Cámara de Diputados del Congreso de la Unión de México.

[3] Diario Oficial de la Federación (2005). *Ley general de las personas con discapacidad*, publicada el 10 de Junio de 2005, modificada el 1 de Agosto de 2008 y abrogada el 30 de Mayo de 2011. Diario Oficial de la Federación, Órgano del Gobierno Constitucional de los Estados Unidos Mexicanos. https://www.dof.gob.mx/nota_detalle.php?codigo=2044351&fecha=10/06/2005#gsc.tab=0.

[4] GobCDMX (2017). *Diccionario de lengua de señas mexicana de la Ciudad de México*. https://pdh.cdmx.gob.mx/storage/app/media/banner/Dic_LSM_2.pdf. INDEPEDI. Ciudad de México.

[5] Cruz-Aldrete, M. (2008). *Gramática de la lengua de señas mexicana*. Tesis de doctorado. El Colegio de México.

[6] Mendoza-Tene (2025). *Desarrollo de un sistema automático de reconocimiento de la LSM para dispositivos móviles*. *Revista Internacional de In-*

vestigación e Innovación Tecnológica. 12(72), 1-23. https://riit.com.mx/apps/site/files_v2450/lsm_jal_1_riit_ene-feb_2025.pdf.

[7] García-Gil, G., López-Armas, G. D. C., Sánchez-Escobar, J. J., Salazar-Torres, B. A., & Rodríguez-Vázquez, A. N. (2024). *Real-Time Machine Learning for Accurate Mexican Sign Language Identification: A Distal Phalanges Approach*. *Technologies*, 12(9), 152. DOI: 10.3390/technologies12090152.

[8] Trujillo-Romero, F., & García-Bautista, G. (2023). *Mexican Sign Language corpus: Towards an automatic translator*. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(8), 1-24. DOI: 10.1145/3591471.

[9] Pichardo-Lagunas, O., & Martínez-Seis, B. C. (2015). *Resource Creation for Automatic Translation System from Texts in Spanish into Mexican Sign Language*. *Res. Comput. Sci.*, 100, 129-137.

[10] Pichardo-Lagunas, O., Partida-Terrón, L., Martínez-Seis, B., Alvear-Gallegos, A., & Serrano-Olea, R. (2016). *Sistema de traducción directa de español a LSM con reglas marcadas*. *Res. Comput. Sci.*, 115, 29-41.

[11] Pichardo-Lagunas, O., Martínez-Seis, B., Ponce-de-León-Chávez, A., Pegueros-Denis, C., Muñoz-Guerrero, R. (2017). *Linguistic Restrictions in Automatic Translation from Written Spanish to Mexican Sign Language*. In: Sidorov, G., Herrera-Alcántara, O. (eds) *Advances in Computational Intelligence. MICAI 2016. Lecture Notes in Computer Science()*, vol 10061. Springer, Cham. DOI: 10.1007/978-3-319-62434-1_8.

[12] Martínez-Seis, B., Pichardo-Lagunas, O., Hernández-Morales, E., Rivera-Rodríguez, O., & Miranda, S. (2025). *Automatic Translation of Sentences to Mexican Sign Language: Rule-based Machine Translation and Animation Synthesis in Avatar*. *Computación y Sistemas*, 29(1), 145-155.

[13] Hernández-Cruz, J. (2019). *Traducción de Texto en Español a Texto LSM Usando Aprendizaje Profundo*. Tesis de maestría. Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo.

[14] Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., & Abid, A. (2021). *Advances in machine translation for sign language: approaches, limitations, and challenges*. *Neural Computing and Applications*, 33(21), 14357-14399. DOI: 10.1007/s00521-021-06079-3.

[15] Lara-Ortiz, V., Fuentes-Aguilar, R. Q., & Chairez, I. (2025). *Spanish to Mexican Sign Language*

glosses corpus for natural language processing tasks. *Scientific Data*, 12(1), 702.

[16] Ríos-Figueroa, H. V., Sánchez-García, A. J., Sosa-Jiménez, C. O., y Solís-González-Cosío, A. L. (2022). Use of spherical and cartesian features for learning and recognition of the static mexican sign language alphabet. *Mathematics*, 10(16), 2904. DOI: 10.3390/math10162904.

[17] Rodríguez, R. F., Rosas, F. J. P., Zuñiga-Madrid, L. Á., & Arguijo, P. (2021). Reconocimiento de las señas estáticas del LSM con características basadas en aprendizaje profundo. *Research in Computing Science*, 150(6), pp. 303-311.

[18] Fregoso, J., Gonzalez, C. I., & Martinez, G. E. (2021). Optimization of convolutional neural networks architectures using PSO for sign language recognition. *Axioms*, 10(3), 139. DOI: 10.3390/axioms10030139.

[19] Varela-Santos, Héctor, Morales-Jiménez, Anahí, Córdova-Esparza, Diana-Margarita, Terven, Juan, Mirelez-Delgado, Flabio Darío, & Orenday-Delgado, Adán. (2021). Assistive Device for the Translation from Mexican Sign Language to Verbal Language. *Computación y Sistemas*, 25(3), 451-464. DOI: 10.13053/cys-25-3-3459.

[20] Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., & Palmer, M. (2000). A machine translation system from English to American Sign Language. In *Envisioning Machine Translation in the Information Future: 4th Conference of the Association for Machine Translation in the Americas*, AMTA 2000 Cuernavaca, Mexico, October 10-14, 2000 Proceedings 4 (pp. 54-67). Springer Berlin Heidelberg. DOI: 10.1007/3-540-39965-8_6.

[21] d'Armond, L. S., & Speers, M. S. (2002). Representation of American sign language for machine translation (Doctoral dissertation, Georgetown University).

[22] Porta, J., López-Colino, F., Tejedor, J., & Colás, J. (2014). A rule-based translation from written Spanish to Spanish Sign Language glosses. *Computer Speech & Language*, 28(3), 788-811. DOI: 10.1016/j.csl.2013.10.003

[23] Kouremenos, D., Ntalianis, K., & Kollias, S. (2018). A novel rule based machine translation scheme from Greek to Greek Sign Language: Production of different types of large corpora and Language Models evaluation. *Computer Speech & Language*, 51, 110-135. DOI: 10.1016/j.csl.2018.04.001.

[24] Luqman, H., & Mahmoud, S. A. (2019). Automatic translation of Arabic text-to-Arabic sign language.

Universal Access in the Information Society, 18(4), 939-951. DOI: 10.1007/s10209-018-0622-8.

[25] Brants, T., Popat, A., Xu, P., Och, F. J., & Dean, J. (2007, June). Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 858-867). <https://aclanthology.org/D07-1090.pdf>.

[26] Zhang, B., Haddow, B., & Birch, A. (2023). Prompting large language model for machine translation: A case study. In *International Conference on Machine Learning* (pp. 41092-41110). PMLR.

[27] Devlin, J. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[28] Yang, Z. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. *arXiv preprint arXiv:1906.08237*.

[29] Liu, Y. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 364.

[30] Cate, H., & Hussain, Z. (2017). Bidirectional american sign language to english translation. *arXiv preprint arXiv:1701.02795*.

[31] Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2020). Text2Sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, 128(4), 891-908. DOI: 10.1007/s11263-019-01281-2.

[32] Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., & Ringel Morris, M. (2019, October). Sign language recognition, generation, and translation: An interdisciplinary perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 16-31). DOI: 10.1145/3308561.3353774.

