



**INSTITUTO POLITÉCNICO NACIONAL**  
UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERÍA Y TECNOLOGÍAS  
AVANZADAS  
**PROTOCOLO**  
**INGENIERÍA TELEMÁTICA**



DTA-PPT-01

TÍTULO DEL PROTOCOLO				
<b>Obtención de reglas de traducción de español a Lengua de Señas Mexicana con redes neuronales artificiales</b>				
DATOS DEL PROTOCOLO				
<b>Número de Revisión</b> <small>(Primera, segunda, tercera o Protocolo para Registro)</small>	<b>Registro</b>	<b>Semestre</b>	<b>23-2</b>	
<b>Número Proyecto Asignado</b> <small>(Número asignado por el profesor de Especialidad)</small>	<b>18</b>	<b>Fecha</b> <small>(Fecha programada)</small>	<b>26 de junio del 2023</b>	
<b>Confidencialidad</b> <small>(Público o confidencial, incluir documento que lo avale)</small>	<b>Público</b>	<b>Número de Hojas</b> <small>(Cantidad de Hojas del Protocolo)</small>	<b>18</b>	
<b>Patrocinador</b> <small>(En caso de existir, incluir el nombre en caso contrario dejar en blanco)</small>				
<b>Número Convenio o Registro</b> <small>(Incluir número de convenio patrocinio o número de proyecto de investigación que patrocina)</small>				
ALUMNO 1				
<b>DATOS ALUMNO 1</b>			<b>FIRMA</b>	
Nombre del Alumno	Gómez García Carlos Eduardo			
Número de boleta	2019640224			
Teléfono	5530412102			
Correo electrónico	cgomez.egarcia@gmail.com			
ALUMNO 2				
<b>DATOS ALUMNO 2</b>			<b>FIRMA</b>	
Nombre del Alumno				
Número de boleta				
Teléfono				
Correo electrónico				
ALUMNO 3				
<b>DATOS ALUMNO 3</b>			<b>FIRMA</b>	
Nombre del Alumno				
Número de boleta				
Teléfono				
Correo electrónico				
<b>DATOS ASESOR 1</b>			<b>VISTO BUENO ASESOR 1</b>	
Nombre Asesor (Grado Académico)	Dra. Bella Citlali Martínez Seis			
Academia	Informática	Interno	<input checked="" type="checkbox"/> Externo <sup>1</sup>	
Cédula Profesional (Obligatorio)	11878306			
Correo electrónico	bcmartinez@ipn.mx			
<b>DATOS ASESOR 2</b>			<b>VISTO BUENO ASESOR 2</b>	
Nombre Asesor (Grado Académico)	Dra. Obdulia Pichardo Lagunas			
Academia	Informática	Interno	<input checked="" type="checkbox"/> Externo	
Cédula Profesional (Obligatorio)	7232281			
Correo electrónico	opichardola@ipn.mx			
<b>DATOS ASESOR 3</b>			<b>VISTO BUENO ASESOR 3</b>	
Nombre Asesor (Grado Académico)				
Academia		Interno	<input type="checkbox"/> Externo	
Cédula Profesional (Obligatorio)				
Correo electrónico				
<b>NOMBRE DEL PROFESOR TITULAR</b>		Sandra Martínez Solís		
<b>NOMBRE DEL PROFESOR DE ESPECIALIDAD</b>		Dr. Noé Torres Cruz		

<sup>1</sup> En caso de Asesores Externos, deberá incluirse copia de su Cédula Profesional y Curriculum Vitae resumido en un archivo anexo al Protocolo.

# INSTITUTO POLITÉCNICO NACIONAL

UNIDAD PROFESIONAL INTERDISCIPLINARIA EN  
INGENIERÍA Y TECNOLOGÍAS AVANZADAS

---

## Obtención de reglas de traducción de español a Lengua de Señas Mexicana con redes neuronales artificiales

---

*Asesores:*

Dra. Bella Citlali Martinez

Seis

Dra. Obdulia Pichardo

Lagunas

*Presentan:*

Carlos Eduardo Gómez

García

### Resumen

En México, 2.4 millones de personas tienen discapacidad auditiva y utilizan la Lengua de Señas Mexicana (LSM). A pesar de que la LSM es una lengua con su propia sintaxis, gramática y léxico, no cuenta con una estandarización y recibe poca atención por parte de la población y el gobierno. En este trabajo se propone desarrollar un sistema con Traducción Automática (TA) basada en reglas del español escrito a LSM considerando la estructura gramatical de cada una de las lenguas. La transformación de la estructura gramatical, del español escrito a LSM, será obtenida con técnicas de aprendizaje automático (*machine learning*), específicamente con algoritmos de redes neuronales artificiales. El conjunto de entrenamiento considerará las estructuras gramaticales obtenidas de oraciones en español obtenidas de transcripciones de videos que con ayuda de un señante de LSM se obtendrán las palabras que son representadas con señas.

**Palabras Clave:** Discapacidad auditiva, Lengua de Señas Mexicana, redes neuronales artificiales, traducción automática.

26 de junio de 2023

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Planteamiento del problema</b>	<b>5</b>
<b>3</b>	<b>Propuesta de solución</b>	<b>7</b>
3.1	Alcances.....	10
<b>4</b>	<b>Objetivos</b>	<b>11</b>
4.1	Objetivo General.....	11
4.2	Objetivos Específicos.....	11
<b>5</b>	<b>Estado del arte</b>	<b>12</b>
<b>6</b>	<b>Marco teórico</b>	<b>20</b>
<b>7</b>	<b>Escenario de pruebas</b>	<b>28</b>
<b>8</b>	<b>Cronograma de actividades</b>	<b>29</b>
	<b>Referencias</b>	<b>35</b>

# Índice de figuras

1	Ejemplo de traducción de español a Español Signado	4
2	Esquema general del sistema propuesto	8

# Índice de tablas

1	Tabla comparativa de proyectos.	16
2	Cronograma de actividades para Proyecto Terminal 1.	29
3	Diagrama de Gantt de actividades para Proyecto Terminal 1.	31

# 1. Introducción

En México, 7.1 millones de personas viven con alguna discapacidad, de este total, el 34 % tiene una discapacidad auditiva, lo que equivale a 2.4 millones, según datos del INEGI. [1]

La lengua de señas es la lengua natural de las personas con discapacidad auditiva. A diferencia del lenguaje oral, la lengua de señas se basa en las expresiones manuales y en diversos movimientos faciales, de los brazos y el cuerpo. Cada país tiene su propia lengua de señas, en México se denomina Lengua de Señas Mexicana (LSM). Cabe mencionar que, debido a la diversidad de lenguas del país, han surgido también lenguas de señas emergentes en diversas comunidades indígenas. [2]

La Lengua de Señas Mexicana (LSM) tiene sus propia sintaxis, gramática y léxico [3]. Sin embargo, su falta de estandarización y la escasa cantidad de estudios sobre ella presentan desafíos para la educación e instrucción de las personas sordas. Además, debido a que su estructura gramatical no es la misma que la del español, es difícil para los usuarios de la LSM comprender y expresar ideas complejas.

En relación a la escasa cantidad de recursos educativos disponibles en LSM, algunos diccionarios han sido desarrollados para ayudar a los usuarios a mejorar su comprensión y expresión en la lengua. Ejemplos de estos recursos incluyen:

- “Manual de Lengua de Señas Mexicana” elaborado por el DIF de Puebla que es un listado de vocabulario en LSM. [4]
- “Manos con voz” que es un diccionario elaborado por la CONAPRED que se conforma por 1113 palabras distribuidas en 15 temas: abecedario, alimentos, animales, antónimos, casa, calendario, colores, escuela, familia, frutas y verduras, números y palabras relacionadas, partes del cuerpo, pronombres-adjetivos-preposiciones-artículos y otras palabras. [5]

Es importante mencionar que estos diccionarios, únicamente muestran en imágenes la secuencia de señas a realizar para decir cierta palabra en español, en ellos no se encuentra explicación alguna de la estructura gramatical en LSM.

Hay otro diccionario elaborado por la Ciudad de México titulado: *Diccionario de Lengua de Señas Mexicana de la Ciudad de México* el cual recopila más de 1000 señas de diversos informantes pertenecientes a comunidades sordas establecidas en la Ciudad de

México. A diferencia de otros diccionarios, este proporciona una explicación detallada de la estructura gramatical y las características de la lengua. Al buscar una palabra en este diccionario, no solo se muestra la secuencia de imágenes, sino que también se incluye información adicional, como se puede observar en la Figura 1. Sin embargo, la estructura gramatical mostrada no corresponde a la de la LSM si no a Español Signado que es una traducción palabra por palabra del español.

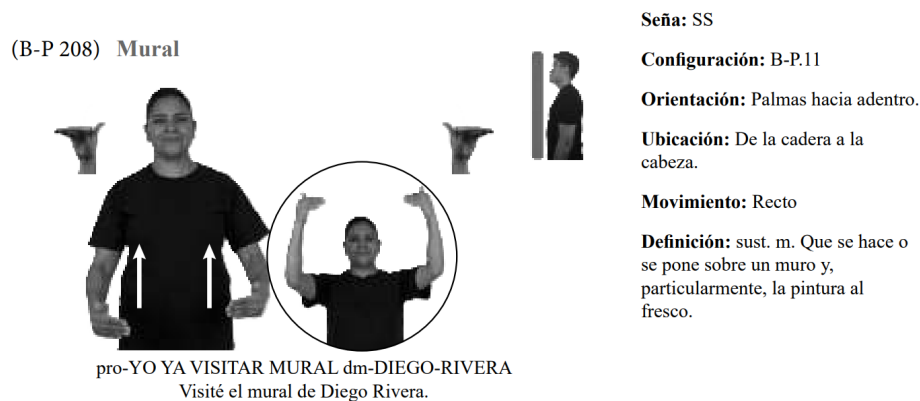


Figura 1: Ejemplo de traducción de español a Español Signado [6]

Existen algunos intérpretes de LSM disponibles, sin embargo, su número es bastante limitado en la actualidad. Al ser una discapacidad que no se ve, recibe muy poca atención de la sociedad y el gobierno [7]. En México solo hay 42 intérpretes certificados de LSM, más otros 200 con algún grado de preparación y unos 150 que no lo son formalmente. [8]

Debido a las grandes diferencias que existen entre el español y la LSM, la gran diversidad de lenguas de señas existentes, la falta de estandarización de la LSM, la dificultad que hay para crear materiales que faciliten su estudio o traducción, además de la falta de intérpretes y de la poca atención que se otorga a las personas con discapacidad auditiva se propone realizar un sistema de traducción automática (TA) de español a LSM considerando la estructura gramatical de cada una de las lenguas utilizando redes neuronales artificiales. Las reglas de transformación gramatical del español a LSM se obtendrán mediante el aprendizaje automático, utilizando ejemplos de transcripciones de lenguaje oral extraídas de videos, donde se transcriben las palabras representadas en LSM con la asistencia de señantes de LSM. Estas reglas de transformación serán empleadas por el sistema de traducción automática del español escrito al lenguaje de señas mediante el uso de procesamiento de lenguaje natural (PLN)

## 2. Planteamiento del problema

La lengua de señas se diferencia del lenguaje oral por estar basada en expresiones faciales y movimientos de manos, brazos y cuerpo. Esta forma de comunicación no solo sirve para expresar ideas y emociones, sino también para transmitir sentimientos abstractos [9]. Sin embargo, uno de los desafíos para su traducción es la confusión entre lo que es gesto y lo que es lingüístico. Los movimientos manuales y corporales que se utilizan en la lengua de señas no son simples gestos, sino que conforman una gramática propia. Una señal puede representar una palabra o varias, y algunas veces es difícil de traducir debido a su uso íntimo. [10]

La Lengua de Señas Mexicana (LSM), como su nombre sugiere, es una lengua en todos los aspectos, ya que se puede apreciar que posee un amplio y preciso vocabulario. Además, sigue ciertas reglas gramaticales. La LSM no se asemeja al español ni en su vocabulario ni en su estructura gramatical, ya que existen señas que no tienen una equivalencia exacta en español e incluso algunas resultan difíciles de traducir. Por lo tanto, las estructuras de las oraciones en español y en la LSM suelen diferir considerablemente, algunos ejemplos son:

- Español: La maestra parece enojada.
- LSM: Maestra mujer (ella) parece enojada.
- Estructura: Sujeto + Verbo + Atributo.
  
- Español: El maestro es enojón.
- LSM: Maestro enojón así.
- Estructura: Sujeto + Atributo + Así.
  
- Español: Los niños ríen.
- LSM: Niños ellos ríen.
- Estructura: Sujeto + Verbo. [11]

Cabe mencionar que la LSM no tiene una estructura gramatical estandarizada, por lo que las estructuras mostradas anteriormente pueden variar dependiendo de la región o de la comunidad sorda donde se encuentre.

Se han realizado diversos trabajos que utilizan técnicas de *Machine Learning* (ML) y Procesamiento de Lenguaje Natural (PLN) para identificar señas de LSM a otros idiomas. Muchos de estos trabajos utilizan Redes Neuronales Convolucionales (CNN) para analizar imágenes, clasificar elementos visuales y realizar tareas de visión artificial [12]. En el contexto de la identificación de señas, las CNN se han utilizado para detectar y clasificar gestos y movimientos, así como para segmentar la imagen de la mano del fondo y extraer características relevantes.

Por otro lado, las Redes Adversarias Generativas (GANs) son utilizadas para generar imágenes sintéticas, lo que permite ampliar el conjunto de datos de entrenamiento ya existente [13]. En el contexto de la traducción de la lengua de señas, se podrían utilizar las GANs para obtener reglas gramaticales de la LSM y así generar traducciones de español a LSM. Esto resultaría muy conveniente, ya que como se ha mencionado en la introducción, existe poco material para su traducción o estudio.

Bajo estas problemáticas se plantea la siguiente pregunta: ¿Cómo obtener de manera automática la estructura gramatical en LSM de una oración escrita en español utilizando redes neuronales artificiales?



### 3. Propuesta de solución

Bajo el problema expuesto anteriormente, se propone realizar un sistema que se divide en dos etapas, la primera consiste en obtener reglas de transformación de español a LSM a través de un modelo de *Machine Learning* y la segunda consiste en utilizar esas reglas generadas para ingresarlas a un sistema de traducción automática (TA).

En primera instancia se recopilarán datos para el entrenamiento del sistema, estos datos se obtendrán de transcripciones de vídeos donde se transcribe a texto lo que se dice oralmente y se transcribe en texto las palabras representadas en LSM con la ayuda de un señante de LSM.

Posteriormente, se realizará un análisis de diferentes algoritmos para obtener las reglas de transformación de español a LSM. Consecuentemente se propondrá un modelo de *Machine Learning* para implementar el o los algoritmos seleccionados para la obtención de las reglas de transformación.

Para el entrenamiento y basados en las evaluaciones de los algoritmos, se propondrán diferentes experimentos con distintos parámetros para evaluar los resultados obtenidos del modelo propuesto. Estos resultados deberán analizarse para realizar los ajustes pertinentes en el modelo propuesto hasta alcanzar los mejores resultados posibles.

Una vez obtenidas las reglas de transformación, estas se utilizarán en el sistema de traducción automática.

Finalmente, se implementará una interfaz gráfica de usuario donde se podrá ingresar una oración en español escrito para así obtener la traducción correspondiente en LSM.

En la Figura 2 se muestra el esquema general del diseño del sistema propuesto.

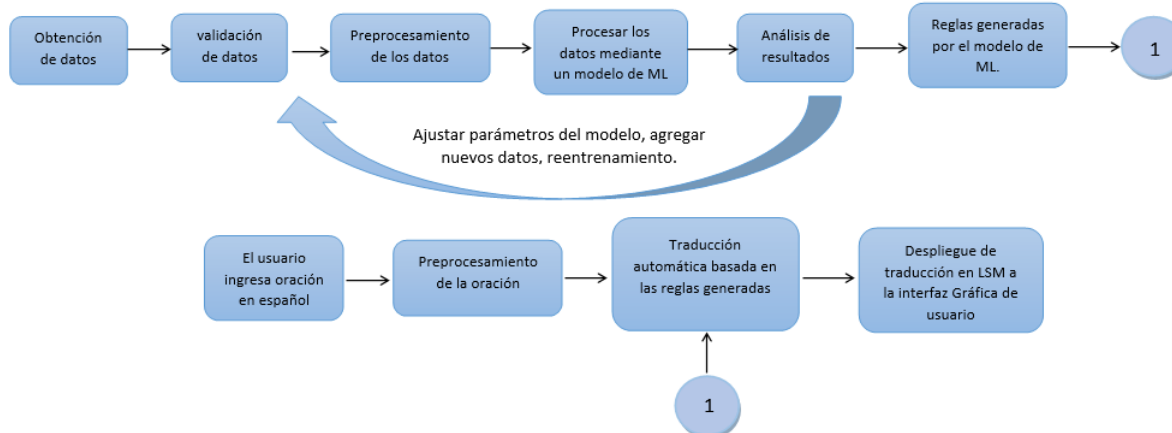


Figura 2: Esquema general del sistema propuesto

## Obtención de datos

Esta etapa implica recopilar los datos que se utilizarán para entrenar al sistema de traducción automática. Esto incluye la recopilación de oraciones en español y con la ayuda de un señante de LSM se obtendrán las palabras que son representadas con señas.

## Validación de datos

En esta etapa con la ayuda de un señante de LSM se validarán los datos obtenidos para así garantizar la fiabilidad de la traducción.

## Preprocesamiento de los datos

En la etapa de preprocesamiento de datos, es importante que las palabras que se introduzcan existan en el corpus recopilado. Si una palabra exacta no está presente, se debe buscar algún sinónimo dentro del corpus para reemplazarla. Además, si hay datos que contienen oraciones demasiado largas, es necesario cortarlas para mejorar la eficiencia del modelo.

Para garantizar que los datos se adapten a la Lengua de Señas Mexicana, se deben quitar algunas palabras que no se utilizan en esta lengua y eliminar acentos o comas. También es recomendable pasar todo el texto a mayúsculas para facilitar el procesamiento.

En resumen, en la etapa de preprocesamiento de datos es fundamental asegurarse de que las palabras sean coherentes con el corpus, recortar oraciones largas, eliminar

palabras innecesarias y normalizar el formato para mejorar la eficiencia del modelo propuesto.

## Procesar los datos mediante un modelo de *Machine Learning*

Esta etapa implica diseñar un modelo de *Machine Learning* (ML) que implemente el o los algoritmo seleccionados para la obtención de reglas de transformación. Para ello, se utilizarán técnicas de *Deep Learning* y se entrenará al modelo con los datos recopilados y validados en las etapa 1 y 2. En el diseño del modelo se propondrán diferentes experimentos, es decir, diferentes parámetros como el número de capas, función de activación a utilizar, número de épocas de entrenamiento, etc. Con el objetivo de realizar diferentes pruebas hasta tener el mejor resultado posible.

## Análisis de resultados

Una vez completados los experimentos propuestos en la etapa anterior, se analizarán los resultados obtenidos a través de métricas como precisión (*precision*) y recuperación (*recall*) para determinar si el modelo de aprendizaje automático necesita ajustes o mejoras adicionales.

## Reglas generadas por el modelo de *Machine Learning*

Una vez analizados los resultados en base a las métricas, se utilizarán las reglas generadas por el modelo para posteriormente ingresarlas al sistema de traducción automática (TA).

## El usuario ingresa oración en español

El usuario deberá ingresar una oración en español a través de una interfaz gráfica de usuario, para que posteriormente se preprocesada antes de ser ingresada al sistema de traducción automática.

## Preprocesamiento de la oración

En esta etapa se realizarán las mismas tareas que en la etapa de preprocesamiento de datos.

## Traducción automática basada en las reglas generadas

En esta etapa se realiza la transformación de español a LSM a partir de las reglas generadas por el modelo de *Machine Learning*.

## Despliegue de traducción en LSM a la interfaz gráfica de usuario

Una vez que el sistema de traducción automática realice la transformación correspondiente de español a LSM, se desplegará en la interfaz gráfica de usuario la traducción generada.

### 3.1. Alcances

- Los datos recopilados serán extraídos de vídeos y de libros aceptados por algunas comunidades sordas.
- Las traducciones que realizará el sistema, estarán acotadas a las estructuras gramaticales y vocabulario obtenido de las fuentes mencionadas anteriormente.
- Las traducciones estarán enfocadas a la LSM del centro, no se contemplan otras variaciones de la LSM.
- Para probar el sistema propuesto se implementará una interfaz gráfica de usuario de escritorio donde se podrá ingresar una oración en español escrito y se obtendrá la traducción en LSM correspondiente.

## 4. Objetivos

### 4.1. Objetivo General

Realizar un sistema de traducción automática de español escrito a LSM considerando la estructura gramatical de cada una de las lenguas utilizando, las reglas de transformación entre las estructuras gramaticales que serán obtenidas con redes neuronales artificiales.

### 4.2. Objetivos Específicos

- Obtener y validar las estructuras gramaticales que se utilizarán para el entrenamiento del sistema.
- Evaluar y seleccionar los algoritmos de *Machine Learning* para la obtención de reglas de transformación de español a LSM.
- Diseñar un modelo de *Machine Learning* que implemente el o los algoritmos seleccionados para la obtención de las reglas de transformación de español a LSM.
- Proponer diferentes experimentos para evaluar el rendimiento del modelo propuesto y analizar los resultados obtenidos en los experimentos para realizar ajustes al modelo en caso de ser necesario.
- Usar técnicas de Procesamiento de Lenguaje Natural para la traducción automática basada en reglas a partir de las reglas obtenidas por el modelo de *Machine Learning*.
- Implementar una interfaz gráfica de usuario que permita interactuar con el sistema de traducción automática de español escrito a Lengua de Señas Mexicana.

## 5. Estado del arte

Se han realizado diferentes estudios para el reconocimiento de lenguaje de señas utilizando inteligencia artificial, debido a su importancia en la comunicación y la inclusión de personas con discapacidades auditivas. En este contexto, el presente estado del arte aborda diferentes investigaciones que han utilizado técnicas de aprendizaje profundo para la detección, reconocimiento y generación de lengua de señas. En particular, se presentan metodologías que han utilizado cámaras de profundidad y redes neuronales recurrentes para reconocer señas de lengua de señas de diferentes países, así como técnicas de aprendizaje no supervisado basadas en redes adversarias generativas para generar vídeos de lengua de señas de alta calidad y realismo. Además, se explica cómo se han evaluado estos modelos en diferentes conjuntos de datos y se muestran los resultados obtenidos mediante diferentes métricas. Finalmente, también se presenta una investigación que utiliza redes neuronales para traducir un lenguaje hablado en vídeos de lengua de señas.

En [14] se presenta un método de reconocimiento dinámico de LSM utilizando una cámara OAK-D. El conjunto de datos utilizado para el entrenamiento, validación y pruebas consiste en 30 señas diferentes, cada una realizada 25 veces por 4 personas diferentes a velocidades distintas, por lo tanto, el conjunto de datos completo consta de 3000 muestras. Se logró una precisión superior al 97 % en el conjunto de pruebas. Se apunta a que en el futuro se pueda extender el número de señas a reconocer e integrar este método en un prototipo.

El trabajo [15] presenta un sistema basado en *Deep Learning* para la detección y reconocimiento de lengua de señas india. El sistema utiliza una red neuronal convolucional (CNN) para la extracción de características de las imágenes de lengua de señas y una red neuronal recurrente (RNN) para reconocer las señas correspondientes. El conjunto de datos utilizados en el sistema contiene 11 palabras, para cada palabra, hay alrededor de 1100 muestras de vídeo. Los autores evaluaron su sistema en distintos conjuntos de datos públicos logrando una precisión del 97 % sobre 11 señas diferentes.

El objetivo de [16] es el de proponer un método novedoso basado en aprendizaje no supervisado para la generación de vídeos de lengua de señas a partir de poses esqueléticas utilizando Redes Adversarias Generativas Dinámicas (*Dynamic GAN*). El método propuesto apunta hacia la generación de vídeos de lengua de seña de alta calidad que son foto realistas y que pueda ser usados en diversas aplicaciones como el reconoci-

to de acciones, análisis del comportamiento humano en público y el monitoreo de las actividades de las personas en entornos concurridos.

El modelo propuesto fue evaluado con 3 conjuntos de datos:

- RWTH-PHOENIX-Weather 2014T: Es un conjunto de datos extraídos de grabaciones de los noticieros “Tagesschau” y “Heute-Journal” de la estación pública alemana “Phoenix” durante un periodo de dos años (2009-2010). En este conjunto de datos se obtuvieron 1980 oraciones en lengua de señas alemana y de las cuales se obtuvo un vocabulario de 911 señas. [17]
- ISL-CSLTR: Consiste en un vocabulario de lengua de señas india de 700 vídeos completamente anotados, 18863 cuadros a nivel de oración y 1036 imágenes a nivel de palabra para 100 oraciones habladas realizadas por 7 signantes diferentes. [18]
- UCF-101: Es un conjunto de datos de reconocimiento de acciones de vídeos de acciones realistas, recolectados de YouTube, teniendo 101 categorías. Estas 101 están agrupadas en 25 grupos, donde cada grupo consiste en 4-7 vídeos de una acción. Los vídeos que son parte del mismo grupo comparten algunas características en común, como el fondo y el similar punto de vista. [19]

Utiliza diferentes métricas para medir los resultados y se aplicaron estas métricas para los 3 conjuntos de datos mencionados anteriormente:

- *Inception Score* (IS): Es una métrica objetiva para evaluar la calidad de las imágenes generadas, específicamente sintéticas producidas por modelos de redes adversarias generativas (GANs) [20]. Para el conjunto de datos PHOENIX, ISL-CSLTR y UCF101 se obtuvieron 8.2, 8.5 y 8.3 respectivamente.
- *Peak Signal-to-Noise Ratio* (PSNR): Compara la calidad de los resultados generados utilizando imágenes de verdad y proporciona la puntuación. El valor PSNR más alto indica una calidad mejorada en los resultados generados. Para el conjunto de datos PHOENIX, ISL-CSLTR y UCF101 se obtuvieron 8.2, 8.5 y 8.3 respectivamente.
- *Fréchet Inception Distance* (FID): Compara la calidad de los resultados generados utilizando imágenes de verdad y proporciona la puntuación. El valor PSNR más alto indica una calidad mejorada en los resultados generados. Para el conjunto de datos PHOENIX, ISL-CSLTR y UCF101 se obtuvieron 8.2, 8.5 y 8.3 respectivamente.

En [21] el objetivo es el de generar la traducción de un lenguaje hablado a partir de vídeos de lengua de señas, teniendo en cuenta los diferentes órdenes de palabras y la gramática. Su conjunto de datos contiene más de .95M cuadros con más de 67 mil señas de un vocabulario de señas de más de 1000 y más de 99 mil palabras de un vocabulario alemán de más de 2.8K.

Las aportaciones de este trabajo se pueden resumir en las siguientes:

- Primera exploración del problema de *Sign Language Translation* (SLT) de vídeo a texto.
- Primer conjunto de datos continuo de SLT disponible públicamente, PHOENIX14T, que contiene segmentos de vídeo, anotaciones de glosas y traducciones de lenguaje hablado.
- El nuevo corpus PHOENIX14T presenta una amplia gama de resultados básicos, que abarcan diversos esquemas de tokenización y atención, así como recomendaciones de parámetros.

El trabajo [22] se centra en el desarrollo de una herramienta para la traducción directa con reglas; la traducción es del español escrito a LSM. Utiliza bases de datos multimedia y procesamiento del lenguaje natural (NLP) para la traducción automática con análisis léxico sintáctico y morfológico.

El léxico utilizado fue obtenido a partir de una investigación realizada para conseguir así un diccionario con la palabra en español y una grabación en vídeo de la seña correspondiente. No se menciona la cantidad de señas con las que está conformado este diccionario, sin embargo, se hace referencia a que se obtuvo un léxico muy ajustado para realizar una primera traducción en el ámbito del hogar.

Resumiendo lo que hace este sistema es, una vez ingresada la frase a traducir, se hace una tokenización y etiquetado de palabras, se clasifican las palabras de acuerdo con su sintaxis y morfología, se realiza un análisis sintáctico para que se cumplan las reglas para la correcta traducción de la frase, se realiza un análisis léxico donde se verifica la existencia de la palabra y por último se obtienen las secuencias de las imágenes que contienen las señas y se crea un vídeo temporal. Finalmente, se muestra en la aplicación el texto ingresado en español, seguido de la traducción en LSM y la secuencia generada en vídeo correspondiente. Se menciona que este software no garantiza la traducción correcta de LSM, pero, es un buen acercamiento para adquirir las bases de esta lengua.

En la tesis [23] se realizó un traductor en español a texto en glosa de LSM utilizando técnicas de *Deep Learning*, con el objetivo de ser una herramienta para el apoyo de traducciones que realiza una plataforma de apoyo educativo.

Para el entrenamiento del modelo se tomaron en cuenta 3 corpus diferentes.

1. El primero de ellos consiste en un repositorio formado con un corpus de aproximadamente 125 idiomas diferentes, todos con oraciones de idioma origen a idioma inglés. Cada uno de estos corpus se conforma por más de 100 mil oraciones.
2. Corpus de español a lengua de señas española (LSE) con 229 oraciones escritas en glosa
3. Corpus conformado por 10 oraciones en LSM escritas en glosa.



En las pruebas realizadas para la traducción a LSM se obtuvo una precisión del 83 % en las métricas, sin embargo, el autor menciona que la traducción no es precisa, pero se obtiene una oración con sentido.

El sistema de traducción utilizado en [24] fue el de reglas, transferencia léxica y sintáctica a través de árboles sintácticos. Para la transformación del español a las estructuras formales e informales de la LSM se construye un árbol a partir de la estructura original en español y sobre este árbol se realiza una búsqueda en orden para obtener la estructura formal en LSM y una búsqueda en post-orden para obtener la estructura informal.

Se diseñó un modelo con validación de estructuras sintácticas básicas y reglas de equivalencia en español. El sistema realiza dos tipos de traducciones: una formal basada en el español signado y otra informal basada en el uso cotidiano del LSM. Se evaluó la reducción dada por las palabras vacías, las estructuras sintácticas, las restricciones léxicas y la corrección de la seña elegida.

El sistema se enfoca en un vocabulario restringido dado por el libro *Manos con Voz*, algunos árboles sintácticos seleccionados y oraciones en tiempo presente. La traducción propuesta utiliza tres corpus principales: el diccionario de señas con palabras en español, un corpus de colocaciones y locuciones de LSM y un corpus de sinónimos.

Se obtuvo un 85.7 % de precisión tanto en traducciones formales e informales. Se comparó con otros sistemas de traducción automática como el HETA y SIGNSLATOR que obtuvieron un 53.9 % y 60.32 % de precisión respectivamente.

Se detectaron complicaciones con el etiquetado. La estructura gramatical sujeto+verbo+adverbio fue la más afectada porque identifica los verbos como sustantivos o adjetivos, lo que hace que los signos mostrados en LSM no correspondan a la semántica de las oraciones. También hay una dificultad en los verbos que actúan como sustantivo en una oración.

En [25] se presentan un enfoque novedoso para la producción automática de lenguaje de señas utilizando la traducción automática neuronal (NMT), redes generativas adversarias (GAN) y la generación de movimiento. El sistema es capaz de producir vídeos de señas a partir de oraciones en lenguaje hablado, requiriendo anotaciones mínimas para el entrenamiento y descomponiendo la tarea en subprocesos dedicados.

El conjunto de datos utilizado en este trabajo consta de 3 señantes que realizan 200 oraciones diferentes, con cada oración que contiene un promedio de 10 palabras.

El sistema desarrollado emplea métodos recientes de traducción automática basados en RNN, a saber, enfoques en *Neural Machine Translation* (NMT) basados en atención, para realizar la traducción de secuencias de lenguaje hablado a secuencias de glosas de la lengua de señas. El sistema utiliza una arquitectura codificador-decodificador y una

red de generación de señas condicionada por pose con una arquitectura codificador-decodificador-discriminador que produce la salida de un vídeo de señas.

Los resultados cualitativos muestran que las secuencias sintéticas generadas por la red se mantienen cerca de su verdad fundamental tanto en términos de movimiento como de apariencia. Sin embargo, todavía hay algunos desafíos que deben abordarse, como mejorar la precisión para las secuencias de datos corruptas y abordar las dificultades para formar brazos y manos completamente en ciertas posiciones.

En la Tabla I se presenta una comparativa de las características técnicas que presentan los trabajos que se mencionaron anteriormente.

Tabla 1: Tabla comparativa de proyectos.

Titulo	Objetivo	Sentido de la traducción	Tamaño del set de datos	Precisión	Técnica utilizada	Puntos de mejora
<i>Automatic Recognition of Mexican Sign Language Using a Depth Camera and Recurrent Neural Networks.</i> [14]	Reconoce la seña ingresada y se obtiene la palabra correspondiente en español escrito.	De Lengua de señas a español signado. (No se realiza una traducción)	3000 muestras de vídeo compuestas por 30 señas diferentes	Mas del 97 % en el conjunto de pruebas	Redes neuronales recurrentes (RNN), LSTM y GRU	Extender el número de señas a reconocer
<i>Deepsign: Sign Language Detection and Recognition Using Deep Learning.</i> [15]	Reconoce la seña ingresada en Lengua de Señas India y se obtiene la palabra correspondiente en hindi	De lengua de señas India a hindi signado (No se realiza una traducción)	1100 muestras de vídeo compuesto por 11 señas diferentes	97 %	Una estructura de RNN usando una sola capa de LSTM seguido de una GRU, GANs Dinamicas	Extender el número de señas a reconocer

<i>Dynamic GAN for High-Quality Sign Language Video Generation from Skeletal poses using Generative Adversarial Networks.</i> [16]	Se enfatiza la traducción de la postura esquelética en vídeos de lengua de señas utilizando GANs	N/A	1980 oraciones en lengua de señas alemana, 100 oraciones en lengua de señas india y 700 vídeos. Estos vídeos muestran acciones divididas en 101 categorías, que se agrupan en 25 grupos.	SSIM: 0.901, 0.937 y 0.925. IS: 8.2, 8.5 y 8.3. PSNR: 8.2, 8.5 y 8.3	GANs Dinámicas	Extender el número de señas a reconocer
<i>Neural Sign Language Translation.</i> [21]	Crear traducciones de lenguaje hablado a lengua de señas utilizando videos, considerando la variación en el orden de las palabras y la gramática	De Alemán a lengua de señas	>.95M de cuadros de video con >67 mil señas de un vocabulario de >1000 señas y >99 mil palabras de un vocabulario alemán de >2.8k	ROUGUE 43.85 usando una GRU, ROUGE 43.85 usando Loung attention, ROUGE 46.02 utilizando un batch size de 1	CNN, diferentes métodos de tokenización, como RNN-HMM y redes de codificador-decodificador basadas en atención	Extender los mecanismos de atención al dominio espacial para alinear los bloques de construcción de las señas.
Sistema de traducción directa de español a LSM con reglas marcadas. [22]	Se centra en el desarrollo de una herramienta para la traducción directa con reglas marcadas	De español escrito a LSM	No se menciona el dato exacto	No se menciona	Bases de datos multimedia, NLP y traducción automática basada en reglas marcadas	No se garantiza la traducción correcta en LSM

<p>Traducción de Texto en Español a Texto LSM usando Aprendizaje Profundo.</p> <p>23</p>	<p>Traductor de texto en español a texto en lengua de señas</p>	<p>De español escrito a LSM</p>	<p>Corpus de más de 125 idiomas distintos al inglés, con alrededor de 100 mil oraciones cada uno. También hay un corpus de español a LSE, el cual consta de 229 oraciones escritas en glosa, y se utilizaron 9 oraciones en español con su respectiva traducción a LSM.</p>	<p>83.94 % de precisión y 0.5946 con la métrica BLEU</p>	<p>Traducción automática basada en reglas y aprendizaje profundo</p>	<p>ampliar el número de estructuras y léxico</p>
<p><i>Linguistic Restrictions in Automatic Translation from Written Spanish to Mexican Sign Language.</i></p> <p>24</p>	<p>Desarrollo de un corpus de signos en LSM, un diccionario de colocaciones inter-lingüísticas entre español y LSM y un diccionario de sinónimos en LSM</p>	<p>De español escrito a LSM</p>	<p>Vocabulario obtenido del libro Manos con voz por 1113 palabras</p>	<p>85.7 % de precisión tanto en traducciones formales e informales</p>	<p>Traducción automática basada en reglas usando árboles sintácticos</p>	<p>dificultades con el etiquetado y ampliar el número de estructuras y el léxico</p>

<i>Text2Sign: Towards Sing Language Production Using Neural Machine Transaltion and Generative Adversarial Networks.</i> [25]	Producir vídeo de señas a partir de oraciones en lenguaje hablado	De texto escrito a lengua de señas	200 oraciones diferentes, cada oración contiene un promedio de 10 palabras	La red generó secuencias sintéticas similares a la verdad en cuanto a movimientos y apariencia	NMT y GANs	Mejorar la precisión de datos corruptos y solucionar problemas en la formación de brazos y manos en algunas posiciones
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	---------------------------------------------	----------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------	---------------	---------------------------------------------------------------------------------------------------------------------------------------------------------

En conclusión, la utilización de técnicas de inteligencia artificial para la detección, reconocimiento y generación de lenguaje de señas ha avanzado significativamente en los últimos años. Los resultados obtenidos en los diferentes conjuntos de datos demuestran que estas técnicas son prometedoras y tienen un gran potencial en diversas aplicaciones. Sin embargo, como se muestra en la Tabla 1, muchos de los trabajos se enfocan solamente en identificar unas pocas señas o en generar vídeos de señas, sin realizar una traducción que considere la estructura gramatical. Aquellos que sí realizan la traducción, suelen tener un conjunto de datos de entrenamiento muy limitado y aún necesitan mejorar en la precisión de los modelos para señas más complejas.

Existen diferencias fundamentales entre el trabajo propuesto y los trabajos [22] y [23]. Ambos enfoques indican de manera explícita reglas gramaticales propuestas para traducir del español a LSM, aunque [23] combina la traducción automática basada en reglas con un modelo de aprendizaje profundo que a través del aprendizaje previo realiza la búsqueda de la mejor combinación para mostrar la traducción. Además, [23] cuenta con un conjunto de entrenamiento de solo 9 pares de oraciones en español a LSM para su modelo de aprendizaje profundo.

Por otro lado, el sistema de traducción automática que se pretende realizar no se basará en la indicación explícita de reglas gramaticales para realiza la traducción. En su lugar, se obtendrán reglas de transformación de la estructura gramatical del español a LSM mediante el aprendizaje automático, mediante ejemplos recopilados de oraciones de transcripciones de videos que posteriormente con ayuda de un señante de LSM se obtendrán las palabras para las señas correspondientes. Estas reglas de transformación serán empleadas por el sistema de traducción automática del español escrito al despliegue de señas utilizando Procesamiento de Lenguaje Natural (PNL).

## 6. Marco teórico

En este capítulo se presentan definiciones, conceptos y teoría, en la cual se respalda el presente trabajo. Se mencionan elementos importantes de la inteligencia artificial y de la Lengua de Señas Mexicana.

### 1. Traducción

La traducción implica el proceso de expresar en un idioma diferente algo que ya ha sido expresado o escrito en otro idioma. El término puede referirse tanto a la interpretación dada a un texto o discurso como al trabajo realizado por el traductor. Hay varios tipos de traducción. Una de ellas es la traducción directa, que se lleva a cabo desde un idioma extranjero al idioma del traductor. Por otro lado, la traducción inversa se realiza desde el idioma del traductor hacia un idioma extranjero. [26]

#### 1.1 Traducción automática

En el contexto de las ciencias de la computación se encuentra lo que se conoce como traducción automática (TA). La traducción automática implica el uso de inteligencia artificial para traducir un texto de un idioma a otro sin la intervención humana. En lugar de simplemente traducir el texto de manera literal, la traducción automática moderna busca comunicar el significado completo del texto original en el idioma de destino. Para lograr esto, analiza todos los elementos del texto y reconoce cómo las palabras se relacionan entre sí.

Existen diferentes enfoques de traducción automática:

- Traducción basada en reglas
- Traducción automática estadística
- Traducción automática híbrida

##### 1.1.2 Traducción Basada en reglas

Para este tipo de traducción se requiere que unos especialistas en idiomas desarrollen reglas lingüísticas y diccionarios para temas o sectores específicos. La traducción automática basada en reglas utiliza estos recursos para traducir con precisión contenidos específicos. El proceso consta de los siguientes pasos:

1. El software de traducción automática analiza el texto de entrada y crea una representación intermedia.
2. Utilizando las reglas gramaticales y los diccionarios como referencia, el software convierte la representación intermedia en la lengua de destino.

### 1.1.3 Traducción automática estadística

Este tipo de traducción a diferencia de la basada en reglas utiliza técnicas de ML para traducir textos. Los algoritmos de ML examinan grandes cantidades de traducciones humanas previas en busca de patrones estadísticos. Luego, cuando se enfrentan a un nuevo texto fuente, el software hace una suposición inteligente sobre cómo traducirlo. Esto se logra haciendo predicciones basadas en la probabilidad estadística de que una palabra o frase específica aparezca junto a otra palabra o frase en el idioma de destino.

### 1.1.4 Traducción automática híbrida

Esta técnica consiste en utilizar dos o más modelos de traducción automática en un solo software. Un enfoque híbrido puede utilizarse para mejorar la eficacia de un modelo de traducción único. Este proceso de traducción automática suele emplear subsistemas de traducción automática basados tanto en reglas como en estadísticas. La traducción final es el resultado de la combinación de las salidas de todos los subsistemas. [27]

## 2. Inteligencia Artificial

La inteligencia artificial, también conocida como IA, se enfoca en resolver problemas cognitivos relacionados con la inteligencia humana, como el aprendizaje, la resolución de problemas y el reconocimiento de patrones. A pesar de que a menudo se asocia con escenarios futuristas o robóticos, la IA es una rama avanzada de la ciencia informática que consta de varias tribus, incluyendo simbolistas, conexionistas, evolutivos, bayesianos y analogistas. Dos subcampos importantes de la IA son el Machine Learning (ML) y el Deep Learning (DL), que se basan en técnicas supervisadas y no supervisadas. En general, la IA aprende más rápido y se vuelve más inteligente a medida que se le proporcionan más datos. [28]

### 2.1 *Machine Learning*

El término aprendizaje automático o *Machine Learning* se refiere a diversas técnicas de aprendizaje y reconocimiento de patrones basadas en la teoría bayesiana. Básicamente, se trata de un conjunto de algoritmos que pueden aprender de datos y hacer predicciones en función de ellos, extraer estructuras de datos y clasificarlos en descripciones concisas. Se utiliza cuando la programación explícita resulta poco práctica o demasiado rígida. En lugar de codificar un programa para producir una salida específica a partir de una entrada determinada, el aprendizaje automático genera un modelo estadístico que produce la "salida adecuada" basada en un patrón reconocido a partir de ejemplos anteriores. La precisión de un modelo de aprendizaje automático depende

de la calidad y cantidad de datos utilizados para entrenarlo. Con los datos adecuados, un modelo de aprendizaje automático puede analizar problemas de gran dimensión con miles de millones de ejemplos para determinar la función óptima capaz de predecir una salida con una entrada determinada. Los modelos de aprendizaje automático también proporcionan confianza estadística en las predicciones y su desempeño general. [28]

### 2.1.2 Procesamiento de Lenguaje Natural (PLN)

El procesamiento de lenguaje natural (PLN) es una técnica de *Machine Learning* que permite a las computadoras entender, procesar y manipular el lenguaje humano. El PLN es esencial para analizar de manera eficiente los datos de texto y voz en profundidad. Puede manejar las variaciones en dialectos, jerga y las irregularidades gramaticales comunes en las conversaciones diarias. [29]

### 2.1.3 Redes Neuronales Recurrentes (RNNs)

Estos algoritmos de *Machine Learning* se utilizan a menudo para problemas que involucran secuencias o datos temporales, como la traducción de idiomas, el reconocimiento de voz y el subtitulado de imágenes. Estos algoritmos utilizan datos de entrenamiento para aprender y se caracterizan por su capacidad para “recordar” información de entradas anteriores y utilizarla en la entrada y salida de datos. A diferencia de las redes neuronales profundas tradicionales, que suponen que los datos de entrada y salida son independientes entre sí, los resultados de las RNNs dependen de elementos anteriores en la secuencia.

Existen arquitecturas variantes de RNN:

- Redes neuronales bidireccionales recurrentes (BRNN)
- Memoria a largo plazo a corto plazo (LSTM)
- Unidades recurrentes cerradas (GRU)

#### 2.1.3.1 Redes neuronales bidireccionales recurrentes (BRNN)

Esta es una versión diferente de la estructura de redes neuronales recurrentes conocida como RNN. Mientras que las RNN unidireccionales solo pueden utilizar información de entrada anterior para realizar predicciones sobre el estado actual, las RNN bidireccionales aprovechan información futura para mejorar su precisión.

#### 2.1.3.2 Memoria a largo plazo a corto plazo (LSTM)

Esta es una estructura de red neuronal recurrente muy popular que fue creada por Sepp Hochreiter y Juergen Schmidhuber como una solución al problema de la desaparición del gradiente. En su documento de investigación *Long Short-Term Memory*, se centraron en el problema de las dependencias a largo plazo, donde si la información



previa que influye en la predicción actual no se encuentra en el pasado reciente, es posible que el modelo RNN no pueda realizar predicciones precisas del estado actual.

Para ilustrar este problema, podemos tomar el ejemplo de la predicción de las palabras en cursiva de la oración “Alice es alérgica a las nueces. Ella no puede comer mantequilla de maní”. Si el contexto de la alergia a las nueces aparece unas pocas oraciones antes, sería difícil para la RNN conectar la información necesaria para realizar una predicción precisa.

Para solucionar este problema, los LSTM (Memorias de Corto y Largo Plazo) tienen “celdas” en las capas ocultas de la red neuronal que tienen tres puertas: una puerta de entrada, una puerta de salida y una puerta de olvido. Estas puertas controlan el flujo de información que se necesita para predecir el resultado en la red. Por ejemplo, si los pronombres de género como “ella” se repitieron varias veces en oraciones anteriores, se puede excluir esta información del estado de celda. [30]

### 2.1.3.3 Unidades recurrentes cerradas (GRU)

Esta versión de RNN se asemeja a los LSTM, ya que también está diseñada para resolver el problema de la memoria a corto plazo de los modelos RNN. Sin embargo, en lugar de usar una información de regulación de “estado de celda”, utiliza estados ocultos y, en lugar de tres puertas, tiene dos: una puerta de reinicio y una puerta de actualización. Al igual que las puertas dentro de los LSTM, las puertas de reinicio y actualización controlan la cantidad y el tipo de información que se retiene.

### 2.1.3.4 Seq2Seq

Esta arquitectura consta de dos redes neuronales recurrentes (RNN). La primera RNN, llamada codificador, se encarga de codificar la secuencia de entrada. La segunda RNN toma el vector codificado y produce la secuencia objetivo.

Esta arquitectura se puede utilizar en distintos escenarios, por ejemplo:

- Resumir: Generar un resumen a partir de un pasaje largo, como una noticia, un capítulo de un libro o un artículo.
- Traducción automática: Generar una frase en otro idioma a partir de una frase en un idioma de origen.
- Responder a una pregunta: Generar una respuesta correcta a una pregunta sobre un texto dado.
- Corrección de errores gramaticales: Generar una frase corregida a partir de una frase que puede contener errores gramaticales u ortográficos. [31]

### 2.1.3.5 *Transformer*

La atención es una técnica que utiliza una capa de pesos para identificar las partes más relevantes de la entrada. Por ejemplo, si el orden es importante en la evaluación de las entradas, la capa de atención aumentará el peso para el orden. La arquitectura del *Transformer* es similar al modelo *Seq2Seq*, pero en lugar de utilizar RNN para retener información entre iteraciones, utiliza una capa que mantiene la posición de cada palabra en una secuencia. [32]

## 3. Redes Adversarias Generativas (GANs)

Las redes adversarias generativas (GANs) son un tipo de red neuronal profunda que se utiliza para generar imágenes sintéticas. La idea y descripción del modelo general llamado GANs se originó en la década de 1990 por Jürgen Schmidhuber en los trabajos *Adversarial Artificial Curiosity* y *Learning Factorial Codes by Predictability Minimization*. En 2014, Ian Goodfellow acuñó el término GANs y popularizó este tipo de modelo a través de su trabajo *Generative Adversarial Nets*.

Para entender GANs, primero se deben comprender los términos generativo y adversarial.

- **Generativo:** Se puede pensar en el término generativo como la capacidad de producir algo. Esto puede ser tomar algunas imágenes de entrada y producir una salida con un toque. Por ejemplo, se puede transformar un caballo en una cebra con cierto grado de precisión. El resultado depende de la entrada y de la calidad del entrenamiento de las capas del modelo generativo para ese caso de uso.
- **Adversario:** Se puede pensar en el término adversario como enfrentar una cosa contra otra cosa. En el contexto de GANs, esto significa enfrentar el resultado generativo (imágenes falsas) contra las imágenes reales presentes en el conjunto de datos. El mecanismo específico se llama discriminador, que implementa un modelo que intenta discriminar entre las imágenes reales y las falsas.

Para explicar aún más y proporcionar un ejemplo de la vida real, Goodfellow hizo una analogía que explica la dinámica presente en los modelos GAN:

“ El modelo generativo se puede pensar como análogo a un equipo de falsificadores, tratando de producir dinero falso y usarlo sin ser detectados, mientras que el modelo discriminativo es análogo a la policía, tratando de detectar el dinero falso. La competencia en este juego impulsa a ambos equipos a mejorar sus métodos hasta que los falsos sean indistinguibles de los verdaderos”.

Goodfellow demostró cómo se puede utilizar el poder computacional moderno para generar ejemplos falsos que parezcan imágenes reales de números, personas, animales y

cualquier cosa que se pueda imaginar. Siempre y cuando se pueda seleccionar cuidadosamente los datos, estos tipos de modelos pueden generar ejemplos novedosos. [33]

## 4. Lengua de Señas

La lengua de señas es una forma de comunicación basada en la producción gestual y la percepción visual. Las lenguas de señas tienen estructuras gramaticales definidas y distintas de las lenguas orales con las que coexisten. Las personas sordas pueden establecer un canal de comunicación con su entorno social utilizando la lengua de señas [34]. Por ejemplo, en México, las personas sordas utilizan la Lengua de Señas Mexicana (LSM), que tiene su propia sintaxis, gramática y léxico. [35]

### 4.1 Lengua de Señas Mexicana

Según el Diario Oficial de la Federación en la Ley General de las Personas con Discapacidad, la Lengua de señas es definida como la lengua de una comunidad de sordos que consiste en signos gestuales articulados con las manos y acompañados de expresiones faciales, mirada intencional y movimiento corporal. Esta lengua tiene función lingüística y forma parte del patrimonio lingüístico de la comunidad de sordos. Además, es tan rica y compleja en gramática y vocabulario como cualquier lengua oral. Al igual que cualquier otra lengua, la lengua de señas varía entre comunidades y tiene la capacidad de crear su propio léxico. Sin embargo, en la actualidad no tiene una forma escrita estandarizada. [6]

Existen distintas variantes de la LSM, que se han desarrollado a lo largo del territorio mexicano debido a la diversidad lingüística y cultural de las comunidades sordas en el país. Algunas de estas variantes son:

- Lengua de Señas Yucateca: es una lengua de señas usada en México y Guatemala, por comunidades mayas. No esta relacionada con la Lengua de Señas Guatemalteca y la Lengua de Señas Mexicana.
- Lengua de Señas de Tijuana: No guarda relación alguna con la Lengua de Señas Mexicana.
- Lengua de Señas chatina: es una lengua de señas emergente de los pueblos chatinos de San Juan Quiahije y Cieneguilla en Oaxaca. Aparentemente no tiene relación con la LSM.
- Lengua de señas de Zinacantán: es una lengua de señas emergente utilizada por un pequeña comunidad Tsotsil en el municipio chiapaneco de Zinacantán. [36]-[39]

### 4.2 Gramática de la LSM

La Lengua de Señas Mexicana (LSM) se estructura utilizando el cuerpo en el espacio. Este espacio tiene tres límites: vertical, horizontal y uno que indica la proximidad de las manos al cuerpo. El límite vertical va desde la cintura hasta la coronilla de la

cabeza y el límite horizontal hasta la altura de los codos con los brazos doblados. Si un movimiento sobrepasa estas dimensiones, se interpretará como una exageración o énfasis. [6]

Aunque no existe una estructura gramatical oficial o estandarizada para la LSM, se ha observado que sigue una de sus estructuras gramaticales es similar a la siguiente:

- Tiempo + Lugar + Sujeto + Objeto + Verbo + Pregunta. Cuando la oración lo requiere se utiliza la negación, pregunta o adverbio.

De manera general, se puede notar que la LSM sigue una estructura cronológica en su ordenamiento de palabras, y no hace uso de artículos ni de género masculino o femenino. La indicación de número singular o plural se logra por medio de señas específicas (como por ejemplo para "hombre", "mujer", "muchos"). Además, algunos gestos faciales son necesarios para comprender el lenguaje. La mayoría de los verbos en LSM requieren del uso de tiempos verbales para poder ser entendidos, y el adverbio (que puede indicar lugar, tiempo, modo, cantidad, afirmación, negación o duda) se coloca junto al verbo, por ejemplo "comer rápido".

La secuencia de las señas en LSM es fundamental para poder entablar una comunicación con una persona sorda. Si se emplea una estructura similar al español hablado, se denominaría "Español Señado" o "Español Signado", pero en ningún caso se estaría utilizando LSM. [40]

#### 4.2.1 Afijos

Según la Real Academia Española (RAE), un afijo es un morfema que se adjunta a una base léxica para modificar su significado o propiedades gramaticales. En la lengua de señas, los afijos no se utilizan de la misma manera que en las lenguas orales. Los prefijos y sufijos se utilizan principalmente en contextos donde predomina el español escrito y son más utilizados por los oyentes que señan.

#### 4.2.2 Prefijos

En la LSM, los morfemas que se utilizan como prefijos son señas que actúan como circunstanciales dependiendo del contexto y a menudo se usan por influencia del español. Estos prefijos fueron utilizados durante muchos años, pero ahora solo quedan algunos. En la (LSM), el morfema que indica el género es la seña de mujer, que se realiza después de las señas que tienen género masculino. El morfema que indica el número se articula antes del nombre y si es significativo, se duplica para hacer énfasis. El morfema que indica el tiempo se utiliza al inicio de la emisión y el señante indica primero el tiempo en el que ocurre su relato. También se puede indicar el tiempo arqueando el tronco hacia atrás para referirse al pasado o hacia delante para ubicar al receptor en el futuro.

### 4.2.3 Sufijos

En la (LSM), los sufijos son morfemas que se colocan después de la seña para dar un significado más específico, a menudo por influencia del español. Anteriormente, los sufijos más utilizados en la LSM eran -ción y -mente. [6]

## 7. Escenario de pruebas

Para realizar las pruebas del sistema de traducción, Se realizará una comparación de la traducción obtenida de la interfaz gráfica de usuario con la que proporcionaría un señante de LSM (con el que ya se tiene contacto). Para ello, se preguntará al señante cómo traduciría la oración en español que se ingresará a la interfaz de usuario y se hará una comparación con la traducción generada.

Este enfoque de comparación permitirá evaluar la calidad de la traducción obtenida en la interfaz gráfica de usuario y mediremos su precisión con la métrica BLEU<sup>1</sup>. Al obtener una retroalimentación del intérprete, podremos detectar posibles errores o problemas en el sistema y mejorar su rendimiento en consecuencia.

---

<sup>1</sup>BLEU (*Bilingual Evaluation Understudy*) es un método de evaluación de la calidad de traducciones realizadas por sistemas de traducción automática. Una traducción tiene mayor calidad cuanto más similar es con respecto a otra referencia, que se supone correcta. BLEU puede calcularse utilizando más de una traducción de referencia. Esto permite una mayor robustez a la medida frente a traducciones libres realizadas por humanos. [41]

## 8. Cronograma de actividades

En la Tabla 2 se presentan las actividades propuestas a realizar durante Proyecto Terminal 1

Tabla 2: Cronograma de actividades para Proyecto Terminal 1.

No.	Nombre de la actividad	Objetivo	Resultados esperados
1	Estudio profundo de algoritmos y soluciones del estado del arte	Conocer los algoritmos y parámetros utilizados para resolver problemas similares	Documento con el estado del arte y marco teórico amplificado.
2	Análisis de requerimientos del sistema	Identificar los elementos necesarios para el funcionamiento del sistema.	Lista de requerimientos funcionales y no funcionales.
3	Investigación de técnicas de pre-procesamiento de datos	Identificar las diversas técnicas de preprocesamiento de datos	Documento que explique las técnicas de preprocesamiento de datos a utilizar.
4	Análisis de librerías de software para ML	Identificar las diferentes herramientas existentes para desarrollar modelos de ML.	Documento que explique la elección de las librerías de software para ML.

5	Análisis de librerías de software para interfaces gráficas de usuario	Identificar las diferentes herramientas existentes para desarrollar interfaces gráficas de usuario.	Documento que explique la elección de la o las librerías de software para interfaces gráficas de usuario.
6	Análisis de algoritmos	Determinar cuales algoritmos con su respectiva arquitectura son los adecuados para utilizarse en el sistema.	Obtener el o los algoritmos mas óptimos para implementarlos en el sistema.
7	Pruebas unitarias de algoritmos de ML	Verificar el correcto funcionamiento de cada algoritmo de ML a utilizar.	Algoritmos de ML funcionando correctamente.
8	Propuesta de diseño preliminar de un modelo de ML	Analizar y organizar el modelo de ML que se implementará en el sistema.	Correcto funcionamiento de los modelos de ML.
9	Diseño UML del sistema	Mostrar visualmente el comportamiento y la estructura del sistema.	Diagramas que representen la estructura del sistema tales como: Casos de Uso, Diagramas de Clases y actividades.
10	Reporte final de PT1	Elaboración del reporte final de PT1	Reporte PT1.



Tabla 3: Diagrama de Gantt de actividades para Proyecto Terminal 1.

Diagrama de Gantt	Semana															
Número de actividad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Actividad 1																
Actividad 2																
Actividad 3																
Actividad 4																
Actividad 5																
Actividad 6																
Actividad 7																
Actividad 8																
Actividad 9																
Actividad 10																

# Referencias

- [1] Universal. «2.4 millones de personas tienen discapacidad auditiva: INEGI». (2021), dirección: <https://universal.org.mx/2-4-millones-de-personas-tienen-discapacidad-auditiva-inegi/>. (accedido el 7 de marzo de 2023).
- [2] Secretaría de Cultura. «Lengua de Señas Mexicana», dirección: <https://mexicana.cultura.gob.mx/es/repositorio/x2abesp3qm-4#:~:text=La%5C%20lengua%5C%20de%5C%20se%5C%C3%5C%B1as%5C%20es,los%5C%20brazos%5C%20y%5C%20el%5C%20cuerpo.%5C&text=La%5C%20LSM%5C%2C%5C%20como%5C%20todo%5C%20lenguaje%5C%2C%5C%20posee%5C%20su%5C%20propia%5C%20gram%5C%C3%5C%A1tica>. (accedido el 7 de marzo de 2023).
- [3] Consejo Nacional para el Desarrollo y la Inclusión de las Personas con Discapacidad. «Lengua de Señas Mexicana (LSM)». (ago. de 2016), dirección: <https://www.gob.mx/conadis/articulos/lengua-de-senas-mexicana-lsm?idiom=es>. (accedido el 7 de marzo de 2023).
- [4] Sistema Municipal DIF Puebla. «Manual de Lengua de Señas Mexicana», dirección: [https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/manual\\_de\\_lengua\\_de\\_senas\\_mexicana-sistema\\_municipal\\_dif\\_puebla.pdf](https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/manual_de_lengua_de_senas_mexicana-sistema_municipal_dif_puebla.pdf). (accedido el 8 de marzo de 2023).
- [5] M. Serafín de Fleischmann Esther y R. González Pérez. «Manos con voz», dirección: [https://www.conapred.org.mx/documentos\\_cedoc/DiccioSenas\\_ManosVoz\\_ACCSS.pdf](https://www.conapred.org.mx/documentos_cedoc/DiccioSenas_ManosVoz_ACCSS.pdf). (accedido el 8 de marzo de 2023).
- [6] C. A. Mercader Flores, L. Escobar Dellamary, M. Pool Westgaard y et al. «Diccionario de Lengua de Señas Mexicana», dirección: [https://pdh.cdmx.gob.mx/storage/app/media/banner/Dic\\_LSM%5C%202.pdf](https://pdh.cdmx.gob.mx/storage/app/media/banner/Dic_LSM%5C%202.pdf). (accedido el 8 de marzo de 2023).
- [7] «¿Cuántos intérpretes de LSM hay en México?», dirección: <https://todorespondio.es/cuantos-interpretes-de-lsm-hay-en-mexico>. (accedido el 9 de marzo de 2023).
- [8] E. Godoy. «Los intérpretes de lengua de señas mexicana intentan afrontar la discriminación por discapacidad», dirección: [https://www.equaltimes.org/los-interpretes-de-lengua-de-senas?lang=en#.ZDMaw-bmJD\\_](https://www.equaltimes.org/los-interpretes-de-lengua-de-senas?lang=en#.ZDMaw-bmJD_). (accedido el 9 de marzo de 2023).

- [9] El Financiero. «Riesgos geológicos», dirección: <https://www.sgm.gob.mx/Web/MuseoVirtual/Riesgos-geologicos/Introduccion-riesgos.html>. (accedido el 9 de marzo de 2023).
- [10] CONICET. «Lenguas de señas: “cada comunidad desarrolló la propia por necesidad»». (ago. de 2012), dirección: <https://www.conicet.gov.ar/lenguas-de-senas-cada-comunidad-desarrollo-la-propia-por-necesidad/>. (accedido el 9 de marzo de 2023).
- [11] M. Martinez, D. Lopez y G. Escobar, “*Tomo 1 Manual de Gramática de la Lengua de Señas Mexicanas*”.
- [12] A. Pathak. «Redes neuronales convolucionales (CNN): una introducción». (ago. de 2022), dirección: <https://geekflare.com/es/convolutional-neural-networks/>. (accedido el 10 de marzo de 2023).
- [13] MathWorks. «Redes generativas antagónicas (GAN)», dirección: <https://la.mathworks.com/discovery/generative-adversarial-networks.html>. (accedido el 10 de marzo de 2023).
- [14] K. Mejía Pérez, D. M. Córdova Esparza, J. Terven, A. M. Herrera Navarro, T. García Ramírez y A. Ramírez Pedraza, «Automatic Recognition of Mexican Sign Language Using a Depth Camera and Recurrent Neural Networks», *Applied Science*, vol. 12, n.º 11, 2022. dirección: <https://doi.org/10.3390/app12115523>.
- [15] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A. B. Gil González y J. M. Corchado, «Deepsign: Sign Language Detection and Recognition Using Deep Learning», *Electronics*, vol. 11, n.º 11, 2022. dirección: <https://doi.org/10.3390/electronics11111780>.
- [16] B. Natarajan y R. Elakkiya, «Dynamic GAN for High-Quality Sign Language Video Generation from Skeletal poses using Generative Adversarial Networks», ago. de 2021. dirección: <https://doi.org/10.21203/rs.3.rs-766083/v1>.
- [17] J. Forster, C. Schmidt, T. Hoyoux et al., «RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus», en *2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Providence, RI, 2012, págs. 17-24.
- [18] R. Elakkiya y B. Natarajan, *ISL-CSLTR: Indian Sign Language Dataset for Continuous Sign Language Translation and Recognition*, Mendeley Data, 2021. DOI: [10.17632/kcmpdxky7p.1](https://doi.org/10.17632/kcmpdxky7p.1). dirección: <https://data.mendeley.com/datasets/kcmpdxky7p/1>.
- [19] A. S. Uluagac, A. k. Roy Chowdhury y V. M. Patel, «Object detection in video using layered object models and scene context», en *2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012, págs. 1-8. DOI: [10.1109/CVPRW.2012.6239247](https://doi.org/10.1109/CVPRW.2012.6239247).
- [20] J. Brownlee. «How to Implement the Inception Score (IS) for Evaluating GANs», dirección: <https://machinelearningmastery.com/how-to-implement-the-inception-score-from-scratch-for-evaluating-generated-images>. (accedido el 27 de marzo de 2023).

- [21] N. C. Camagoz, S. Hadfield, O. Koller, H. Ney y R. Bowden, «Neural Sign Language Translation», en *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 2018, págs. 7784-7793. DOI: [10.1109/CVPR.2018.00817](https://doi.org/10.1109/CVPR.2018.00817).
- [22] P. L. Obdulia, L. Partida Terrón, B. Martínez Seis, A. Alvear Gallegos y R. Serrano Olea, «Sistema de traducción directa de español a LSM con reglas marcadas», Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, dic. de 2016. DOI: [115.29-41.10.13053/rcs-115-1-3](https://doi.org/10.1155/2016/1152941).
- [23] J. C. Hernández Cruz, «Sistema de traducción automática de lenguaje natural a lenguaje de señas mexicano utilizando aprendizaje profundo», Tesis de Licenciatura, Instituto Tecnológico de Orizaba, Orizaba, México, 2019.
- [24] O. Pichardo Lagunas, B. Martínez Seis, A. Ponce de León Chávez, C. Pegueros Denis y R. Muñoz Guerra, «Linguistic Restrictions in Automatic Translation from Written Spanish to Mexican Sign Language», en *Advances in Computational Intelligence. MICAI 2016. Lecture Notes in Computer Science()*, G. Sidrov y O. Herrera Alcántara, eds., vol. 10061, Springer, Cham. DOI: [https://doi.org/10.1007/978-3-319-62434-1\\_8](https://doi.org/10.1007/978-3-319-62434-1_8).
- [25] S. Stoll, N. C. Camgoz, S. Hadfield y et al, «Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks», *Int J Comput Vis*, vol. 128, págs. 891-908, (2020). DOI: <https://doi.org/10.1007/s11263-019-01281-2>.
- [26] J. Pérez Porto y M. Merino, *Traducción - Qué es, definición y concepto*, [En línea]. dirección: <https://definicion.de/traduccion/>, (accedido el 27 de marzo de 2023).
- [27] Amazon Web Services, *¿En qué consiste la traducción automática?*, [En línea]. dirección: <https://aws.amazon.com/es/what-is/machine-translation/>, (accedido el 27 de marzo de 2023).
- [28] Amazon Web Services, *What is Artificial Intelligence (AI)?*, [En línea]. dirección: <https://aws.amazon.com/machine-learning/what-is-ai/>, (accedido el 27 de marzo de 2023).
- [29] Amazon Web Services, *¿Qué es el procesamiento de lenguaje natural (NLP)?*, [En línea]. dirección: <https://aws.amazon.com/es/what-is/nlp/>, (accedido el 27 de marzo de 2023).
- [30] IBM, *Redes neuronales recurrentes*, [En línea]. dirección: <https://www.ibm.com/mx-es/topics/recurrent-neural-networks>, (accedido el 27 de marzo de 2023).
- [31] Ibidem Group, *Traducción automática neuronal Seq2Seq + Transformers*, [En línea]. dirección: <https://www.ibidemgroup.com/edu/traduccion-automatica-neuronal-seq2seq-transformers/>, (accedido el 27 de marzo de 2023).
- [32] F. Sanz, *Transformer: la tecnología que domina el mundo*, [En línea]. dirección: <https://www.themachinelearners.com/transformer/>, (accedido el 17 de abril de 2023).

- [33] C. Hansen, "*Generative adversarial networks explained*", [En línea], jul. de 2022. dirección: <https://developer.ibm.com/articles/generative-adversarial-networks-explained/>, (accedido el 18 de abril de 2023).
- [34] Wikipedia, "*Lengua de señas*", [En línea]. dirección: [https://es.wikipedia.org/wiki/Lengua\\_de\\_se%C3%B1as](https://es.wikipedia.org/wiki/Lengua_de_se%C3%B1as), (accedido el 28 de marzo de 2023).
- [35] CONADIS, "*Lengua de señas mexicana (LSM)*", [En línea]. dirección: <https://www.gob.mx/conadis/articulos/lengua-de-senas-mexicana-lsm?idiom=es>, (accedido el 28 de marzo de 2023).
- [36] Wikipedia, "*Lengua de Señas Yucateca*", [En línea]. dirección: [https://es.wikipedia.org/wiki/Lengua\\_de\\_se%C3%B1as\\_yucateca](https://es.wikipedia.org/wiki/Lengua_de_se%C3%B1as_yucateca), (accedido el 18 de abril de 2023).
- [37] Wikipedia, "*Lengua de señas de Tijuana*", [En línea]. dirección: [https://es.wikipedia.org/wiki/Lengua\\_de\\_se%C3%B1as\\_de\\_Tijuana](https://es.wikipedia.org/wiki/Lengua_de_se%C3%B1as_de_Tijuana), (accedido el 18 de abril de 2023).
- [38] Wikipedia, "*Lengua de Señas Chatina*", [En línea]. dirección: [https://es.wikipedia.org/wiki/Lengua\\_de\\_se%C3%B1as\\_chatina](https://es.wikipedia.org/wiki/Lengua_de_se%C3%B1as_chatina), (accedido el 18 de abril de 2023).
- [39] Wikipedia, "*Lengua de señas de Zinacantán*", [En línea]. dirección: [https://es.wikipedia.org/wiki/Lengua\\_de\\_se%C3%B1as\\_de\\_Zinacant%C3%A1n](https://es.wikipedia.org/wiki/Lengua_de_se%C3%B1as_de_Zinacant%C3%A1n), (accedido el 18 de abril de 2023).
- [40] J. Palacios. «¿Como hablar en lenguaje de señas? (Tercera parte) - Orden y formación de oraciones». (mayo de 2015), dirección: <https://compartirsignos.blogspot.com/2015/05/como-hablar-en-lenguaje-de-senas.html>. (accedido el 10 de marzo de 2023).
- [41] Wikipedia, "*BLEU*", [En línea]. dirección: <https://es.wikipedia.org/wiki/BLUE>, (accedido el 25 de Mayo de 2023).