



POTENCIAR CON IA LA ASISTENCIA VIRTUAL DE USUARIOS DE LENGUAJE DE SEÑAS: REVISIÓN SISTEMÁTICA DE LITERATURA

Christian Hernán Obando Ibarra¹
ORCID: 0000-0003-2326-8934

Richard De Jesús Gil Herrera²
ORCID: 0000-0003-4481-7808

¹ Doctorando, Universidad Americana de Europa (UNADE). Docente de la Universidad de Antioquia (UdeA). Correo: chrisobib@gmail.com

² Profesor-investigador, Universidad Americana de Europa (UNADE). PhD Computer Science & Information Technology. Correo: richard.dejesus@aulagrupo.es

RESUMEN

La traducción automática del lenguaje de señas hacia texto o voz, y su conversión inversa, representó un desafío tecnológico complejo para la accesibilidad comunicativa. Aunque la inteligencia artificial (IA) avanzó de forma significativa durante la última década, la mayoría de las soluciones desarrolladas se concentraron en tareas unidireccionales, lo que limitó la interacción fluida entre personas sordas y oyentes. Este estudio tuvo como objetivo analizar la literatura científica sobre sistemas de IA aplicados al reconocimiento, traducción y producción automática del lenguaje de señas entre 2018 y 2025, con énfasis en los enfoques bidireccionales. Se desarrolló una revisión sistemática siguiendo el modelo de Kitchenham y los lineamientos PRISMA, adaptados a la ingeniería y las ciencias computacionales. La búsqueda se realizó en Scopus, ScienceDirect, Google Scholar y Semantic Scholar, identificando inicialmente 174 estudios, de los cuales 50 fueron seleccionados tras aplicar criterios de inclusión, exclusión y evaluación de calidad metodológica. Los resultados mostraron una transición dominante hacia arquitecturas Transformer (42%), frente a modelos tradicionales CNN-LSTM (18%), enfoques multimodales visión-lenguaje (16%), modelos generativos como GAN y difusión (12%), arquitecturas basadas en grandes modelos de lenguaje (8%) y otros enfoques híbridos (4%). Los sistemas unidireccionales alcanzaron una madurez considerable, mientras que los bidireccionales permanecieron limitados por la escasez de datasets multilingües, la falta de métricas estandarizadas y la escasa validación en contextos reales. Se concluyó que el desarrollo de sistemas de traducción verdaderamente bidireccionales constituyó el reto tecnológico más relevante para lograr una accesibilidad comunicativa plena entre comunidades sordas y oyentes.

Palabras clave: Accesibilidad, Inteligencia Artificial, Lenguaje De Señas.

ABSTRACT

Automatic translation between sign language and text or speech, and its inverse process, represented a complex technological challenge for the communicative accessibility of deaf individuals. Although artificial intelligence (AI) advanced significantly over the past decade, most existing solutions focused on unidirectional tasks, which limited fluent interaction between deaf and hearing individuals. This study aimed to analyze the scientific literature on AI-based systems applied to the recognition, translation, and production of sign language between 2018 and 2025, with an emphasis on bidirectional approaches. A systematic literature review was conducted following Kitchenham's model and the PRISMA guidelines, adapted to engineering and computational sciences. The search was carried out across Scopus, ScienceDirect, Google Scholar, and Semantic Scholar, initially identifying 174 studies, of which 50 were selected after applying inclusion, exclusion, and methodological quality criteria. The results showed a dominant transition toward Transformer-based architectures (42%), compared to traditional CNN-LSTM models (18%), multimodal vision-language approaches (16%), generative models such as GANs and diffusion (12%), large language model-based architectures (8%), and other hybrid approaches (4%). Unidirectional systems reached a considerable level of maturity, while bidirectional systems remained limited by the scarcity of multilingual datasets, the lack of standardized metrics, and limited validation in real-world contexts. It was concluded that the development of truly bidirectional translation systems constituted the most relevant technological challenge for achieving full communicative accessibility between deaf and hearing communities.

Keywords: Accessibility, Artificial Intelligence, Sign Language.

INTRODUCCIÓN

La comunicación humana constituye un proceso multimodal que integra componentes lingüísticos, gestuales y contextuales. Para las personas sordas, el lenguaje de señas representa un sistema lingüístico completo, con estructura gramatical propia y organización visual-espacial, cuyo procesamiento computacional plantea desafíos significativos para las tecnologías de visión por computador y de procesamiento del lenguaje natural.

La mayoría de los desarrollos tecnológicos se concentran en el reconocimiento automático del lenguaje de señas (Sign Language Recognition, SLR) y en su traducción hacia texto o voz (Sign Language Translation, SLT), dejando en segundo plano la generación automática de señas a partir de texto o habla (Sign Language Production, SLP). Esta limitación reduce la comunicación a un solo sentido: permite que una persona oyente comprenda una seña, pero no que una persona sorda reciba una respuesta en su propio idioma. Una comunicación verdaderamente inclusiva requiere sistemas bidireccionales capaces de integrar reconocimiento, traducción y producción dentro de una misma arquitectura.

El surgimiento del aprendizaje profundo, y en particular de las arquitecturas Transformer, transforma el panorama de investigación en este campo, al permitir modelar dependencias temporales complejas y procesar información multimodal. Paralelamente, los modelos generativos avanzados (GAN, difusión, LLM) abren nuevas posibilidades para la producción automática de señas con mayor naturalidad gestual.

A diferencia de trabajos de referencia como el de Bragg et al. (2019), previo a la consolidación del Transformer, esta revisión aporta tres elementos diferenciales: un protocolo sistemático y replicable (Kitchenham, 2007; PRISMA, Page et al., 2021); cobertura del periodo 2018-2025, que captura la transición hacia modelos Transformer, multimodales y generativos; y la diferenciación explícita entre sistemas unidireccionales y bidireccionales, clave para identificar las brechas tecnológicas pendientes.

Sin embargo, el desarrollo de sistemas bidireccionales continúa siendo limitado debido a la escasez de corpus paralelos de gran escala, la diversidad lingüística entre lenguas de señas y la dificultad de modelar computacionalmente las expresiones faciales y corporales que forman parte del significado lingüístico. Esta brecha tecnológica afecta directamente la autonomía de las personas sordas en contextos educativos, de salud y de trámites administrativos, donde la comunicación en tiempo real con personas oyentes resulta fundamental.

Por lo anterior, este estudio tiene como objetivo general analizar críticamente la literatura científica reciente sobre sistemas de inteligencia artificial aplicados al reconocimiento, traducción y producción automática del lenguaje de señas, con énfasis en los enfoques bidireccionales. De forma específica, se busca: a) identificar las tecnologías de IA aplicadas durante el periodo 2018-2025; b) caracterizar sus contribuciones técnicas en términos de precisión, accesibilidad y aplicabilidad; y c) determinar los principales desafíos técnicos y metodológicos que limitan el desarrollo de sistemas bidireccionales.

En consecuencia, la revisión se orienta a responder las siguientes preguntas de investigación: ¿qué tecnologías de IA se han aplicado al reconocimiento, traducción y producción del lenguaje de señas entre 2018 y 2025?; ¿qué contribuciones técnicas presentan estos sistemas en términos de precisión, accesibilidad y aplicabilidad práctica?; y ¿cuáles son los principales desafíos técnicos y metodológicos que persisten para el desarrollo de sistemas plenamente bidireccionales?

ESTADO DEL ARTE

Durante la última década, la investigación en accesibilidad comunicativa evoluciona desde el reconocimiento automático de señas aisladas mediante CNN hacia modelos secuenciales CNN-LSTM capaces de procesar señas continuas en video (Camgöz et al., 2018; Pigou et al., 2018), y a partir de 2019, hacia arquitecturas Transformer que redefinen el estado de la técnica al modelar dependencias temporales

de largo alcance sin mecanismos recurrentes (Camgöz et al., 2020), combinándose progresivamente con representaciones de pose corporal y redes de convolución en grafos (GCN) para mejorar la robustez frente a variaciones visuales (Yan et al., 2018; Rastgoo et al., 2021).

En paralelo, emergen los primeros sistemas de producción automática de lenguaje de señas (SLP), inicialmente basados en GAN capaces de sintetizar video realista a partir de texto (Stoll et al., 2020; Saunders et al., 2020), que evolucionan hacia arquitecturas Transformer para la generación continua en 3D (Saunders et al., 2022) y hacia modelos de difusión. Iniciativas como el proyecto europeo SignON (Vandeghinste et al., 2023) y los talleres WMT-SLT (Müller et al., 2022, 2023) constituyen los primeros esfuerzos por establecer benchmarks e infraestructuras bidireccionales, aunque estos desarrollos continúan siendo minoritarios frente a la producción académica centrada en tareas unidireccionales.

MARCO TEÓRICO

El procesamiento automático del lenguaje de señas se organiza convencionalmente en tres tareas complementarias (Bragg et al., 2019): el reconocimiento del lenguaje de señas (Sign Language Recognition, SLR), que identifica configuraciones manuales y gestuales a partir de video; la traducción del lenguaje de señas (Sign Language Translation, SLT), que convierte dichas secuencias en texto o voz; y la producción del lenguaje de señas (Sign Language Production, SLP), que genera secuencias gestuales a partir de texto o habla. La integración de estas tres tareas dentro de un mismo sistema constituye la base conceptual de la bidireccionalidad comunicativa.

Las arquitecturas Transformer, basadas en mecanismos de atención, permiten modelar relaciones de dependencia entre elementos distantes de una secuencia sin recurrir a estructuras recurrentes (Camgöz et al., 2020), y suelen combinarse con el mecanismo CTC (Connectionist Temporal Classification) para alinear secuencias de longitud variable sin anotaciones manuales, reduciendo la dependencia de glosas intermedias.

Por su parte, los enfoques multimodales visión-lenguaje integran representaciones visuales (video, pose corporal, esqueleto) con embeddings lingüísticos dentro de un mismo espacio semántico, mejorando la alineación entre secuencias gestuales y sus equivalentes textuales.

En el ámbito generativo, las GAN y los modelos de difusión permiten sintetizar secuencias de señas realistas a partir de texto (Guo et al., 2024), mientras que los LLM incorporan capacidades de razonamiento semántico que favorecen la traducción sin glosas (gloss-free translation), es decir, la correspondencia directa entre video y texto sin representaciones intermedias.

Finalmente, en el marco de este estudio la bidireccionalidad se entiende como la capacidad de un sistema para integrar SLR, SLT y SLP dentro de una misma arquitectura de interacción, permitiendo que una persona sorda reciba y emita información en su propio idioma dentro de un mismo entorno comunicativo.

METODOLOGÍA

Este estudio desarrolló una revisión sistemática de literatura siguiendo el enfoque de Kitchenham (2007)³, complementado con los lineamientos estructurales del marco PRISMA (Page et al., 2021) adaptado a la ingeniería y las ciencias computacionales. El proceso se organizó en cinco fases: definición del protocolo de revisión, formulación de preguntas de investigación, estrategia de búsqueda bibliográfica, selección y evaluación de calidad de los estudios, y extracción y síntesis de los datos.

³ Aunque esta publicación tiene más de una década de antigüedad, se mantiene como el marco metodológico de referencia predominante para la realización de revisiones sistemáticas de literatura en ingeniería de software y disciplinas computacionales afines, razón por la cual su uso continúa siendo pertinente y ampliamente validado por la comunidad científica.

La revisión se guió por tres preguntas de investigación: PI1, sobre las tecnologías de IA aplicadas al reconocimiento, traducción y producción del lenguaje de señas entre 2018 y 2025; PI2, sobre las contribuciones técnicas de estos sistemas; y PI3, sobre los desafíos técnicos y metodológicos que persistieron para el desarrollo de sistemas bidireccionales.

La búsqueda se realizó entre enero y febrero de 2025 en las bases de datos Scopus, ScienceDirect, Google Scholar y Semantic Scholar, utilizando la ecuación: (“sign language translation” OR “sign language recognition” OR “sign language production”) AND (“artificial intelligence” OR “deep learning” OR “neural networks”) AND (“translation” OR “generation” OR “bidirectional”).

Tabla 1

Crterios de inclusión y exclusión

Tipo	Crterios
Inclusión	Artículos publicados entre 2018 y 2025, en inglés o español, centrados en el uso de IA aplicada al reconocimiento, traducción o producción del lenguaje de señas, con descripción técnica del modelo y resultados experimentales o validación metodológica.
Exclusión	Artículos duplicados, estudios sin componente tecnológico, investigaciones centradas exclusivamente en análisis lingüístico sin uso de IA, publicaciones sin evaluación experimental o sin acceso al texto completo.

Los estudios que cumplieron los criterios de inclusión se sometieron a una evaluación de calidad metodológica basada en seis criterios binarios (claridad del objetivo, descripción del dataset, detalle de la arquitectura, evaluación experimental cuantitativa, uso de métricas apropiadas y discusión de limitaciones), con un puntaje mínimo de aprobación de 4 sobre 6 puntos.

Con el fin de reducir la subjetividad en el proceso de clasificación, la categorización de las arquitecturas de inteligencia artificial se realizó aplicando los criterios predefinidos del baremo de evaluación de calidad metodológica, evitando juicios discrecionales no sustentados en los criterios establecidos. Adicionalmente, la clasificación de los estudios y el análisis comparativo resultante fueron revisados por el coautor, contrastando la consistencia de las categorías asignadas frente a la evidencia reportada en cada estudio.

La búsqueda inicial arrojó 174 registros, de los cuales se eliminaron 41 duplicados, resultando en 133 estudios. La revisión de títulos y resúmenes excluyó 67 estudios por ausencia de componentes de IA o por enfoque exclusivamente lingüístico, resultando en 66 artículos evaluados a texto completo. Tras aplicar la evaluación de calidad metodológica, se excluyeron 16 estudios adicionales, quedando 50 estudios incluidos en la revisión sistemática final.

Para cada estudio incluido se extrajeron las siguientes variables: tipo de tarea (reconocimiento, traducción o producción), dirección del sistema (Sign→Text, Text→Sign o bidireccional), arquitectura de IA utilizada, dataset empleado, lengua de señas considerada, métricas de evaluación y principales limitaciones reportadas. La información se consolidó en una matriz comparativa que permitió el análisis presentado en la siguiente sección.

RESULTADOS

De los 50 estudios incluidos en la revisión, 30 corresponden a sistemas unidireccionales (reconocimiento, traducción Sign→Text o producción Text→Sign) y 20 a sistemas orientados a la bidireccionalidad. La Tabla 2 presenta una muestra representativa de los estudios analizados, organizada según el tipo de sistema, la arquitectura de IA empleada y el enfoque tecnológico.

Tabla 2

Muestra representativa de estudios analizados (2018-2025)

N°	Estudio	Tipo de sistema	Arquitectura de IA	Enfoque
1	Camgöz et al. (2018)	Unidireccional (Sign→Text)	CNN + LSTM	Traducción neuronal pionera
2	Camgöz et al. (2020)	Unidireccional (Sign→Text)	Transformer + CTC	Reconocimiento y traducción end-to-end
3	Zhou et al. (2021)	Unidireccional (Sign→Text)	Transformer encoder-decoder	Back-translation
4	Yin et al. (2021)	Unidireccional (Sign→Text)	Transformer	Traducción sin glosas
5	Wong et al. (2024)	Unidireccional (Sign→Text)	Vision encoder + LLM	Sign2GPT
6	Yan et al. (2018)	Unidireccional (SLR)	GCN	Modelado espacial esquelético
7	Rastgoo et al. (2021)	Unidireccional (SLR)	CNN + Pose	Reconocimiento basado en pose
8	Stoll et al. (2020)	Unidireccional (Text→Sign)	NMT + GAN	Producción gestual
9	Saunders et al. (2020)	Unidireccional (Text→Sign)	Pose GAN	Video realista
10	Saunders et al. (2022)	Unidireccional (Text→Sign)	Transformer + MDN	Coarticulación a gran escala
11	Duarte et al. (2021)	Unidireccional (SLR/SLT/SLP)	Dataset multimodal	Corpus How2Sign
12	Müller et al. (2022)	Bidireccional (Sign↔Text)	Transformer multimodal	Benchmark WMT-SLT22
13	Vandeghinste et al. (2023)	Bidireccional (Sign↔Speech/Texto)	MT + Avatar	Plataforma SignON
14	Fang et al. (2024)	Bidireccional (Texto↔Sign)	LLM multimodal	SignLLM
15	Zhang et al. (2024)	Bidireccional (Sign↔Text)	Transformer multimodal	Framework bidireccional completo

Nota. Elaboración propia a partir de la matriz de extracción de datos de los 50 estudios revisados.

En términos de arquitectura, la Tabla 3 resume la distribución de las técnicas de inteligencia artificial identificadas en el conjunto completo de 50 estudios. Los modelos basados en Transformer predominan con el 42% de los trabajos, seguidos de los modelos tradicionales CNN-LSTM con el 18%, los enfoques multimodales visión-lenguaje con el 16%, los modelos generativos GAN y de difusión con el 12%, las arquitecturas basadas en grandes modelos de lenguaje con el 8%, y otros enfoques híbridos con el 4%. Los enfoques recientes basados en LLM y en la reducción de la densidad representacional (Wong et al., 2024; Ye et al., 2024) evidencian la transición hacia la traducción sin glosas.

Tabla 3

Distribución de arquitecturas de IA identificadas (n = 50)

Arquitectura	N° de estudios	Porcentaje
Transformer	21	42%

Arquitectura	N° de estudios	Porcentaje
CNN-LSTM	9	18%
Multimodal visión-lenguaje	8	16%
GAN / Difusión	6	12%
Grandes modelos de lenguaje (LLM)	4	8%
Otros enfoques híbridos	2	4%

Respecto a los datasets, la Tabla 4 compara los principales corpus utilizados en los estudios analizados. PHOENIX-2014T se emplea en la mayoría de los trabajos centrados en lengua de señas alemana, mientras que How2Sign constituye el corpus multimodal más extenso disponible para lengua de señas americana. JWSign (Gueuwou et al., 2023) representa el corpus multilingüe de mayor cobertura, al incluir 98 lenguas de señas, aunque con calidad heterogénea entre ellas.

Tabla 4

Comparación de los principales datasets utilizados

Dataset	Lengua de señas	Tamaño aproximado	Limitación principal
PHOENIX-2014T	DGS (alemana)	8 257 oraciones	Dominio restringido (noticias del tiempo)
How2Sign	ASL (americana)	35 000 oraciones	Alta variabilidad, difícil alineación
JWSign	Múltiples (98)	2 530 horas	Calidad heterogénea entre lenguas
CSL-Daily	CSL (china)	20 654 oraciones	Limitado a contexto cotidiano chino

En cuanto a la dirección del sistema, 21 estudios abordan tareas Sign→Text, 9 estudios corresponden a producción Text→Sign, y 20 estudios desarrollan arquitecturas bidireccionales (Sign↔Text o Speech↔Sign), entre los que se destacan los talleres WMT-SLT (Müller et al., 2022, 2023), el proyecto SignON (Vandeghinste et al., 2023) y el marco SignLLM (Fang et al., 2024).

DISCUSIÓN

Los resultados evidencian una transición tecnológica clara desde arquitecturas CNN-LSTM hacia modelos Transformer, consistente con lo reportado en revisiones previas sobre procesamiento de lenguaje natural y visión por computador. Esta transición se explica por la capacidad de los mecanismos de atención para modelar dependencias temporales de largo alcance sin las limitaciones de las estructuras recurrentes, lo que mejora directamente la precisión en tareas de reconocimiento y traducción continua.

Para comprender el alcance real de esta transición resulta necesario contrastar ventajas, limitaciones y condiciones de aplicación de cada familia de arquitecturas. Los modelos CNN, usados en los primeros trabajos del área (Pigou et al., 2018), son apropiados para el reconocimiento de señas aisladas o vocabularios reducidos, pero no modelan por sí solos la dimensión temporal del movimiento gestual, indispensable para señas continuas. La incorporación de módulos LSTM (Rastgoo et al., 2021; Yan et al., 2018) mejoró el reconocimiento de secuencias completas, pero conserva limitaciones estructurales: el procesamiento secuencial impide paralelizar el entrenamiento, y el desvanecimiento del gradiente restringe el modelado de dependencias de largo alcance.

La adopción del Transformer (Camgöz et al., 2020) no es una mejora incremental, sino un cambio cualitativo: el mecanismo de autoatención relaciona directamente cualquier par de posiciones de la

secuencia sin acumular el error del procesamiento paso a paso, habilitando la alineación semántica sin glosas (Ye et al., 2024; Yin et al., 2021) y la integración con modelos de lenguaje preentrenados (Fang et al., 2024; Wong et al., 2024). Esta diferencia de naturaleza —no solo de grado— justifica hablar de un cambio de paradigma, aunque exige grandes volúmenes de datos anotados, escasos para lenguas de señas distintas del ASL o la lengua de señas alemana.

Un patrón similar se observa en la producción automática de señas: las GAN (Saunders et al., 2020; Stoll et al., 2020) y los modelos de difusión (Guo et al., 2024) mejoran el realismo visual, pero ninguna familia modela simultáneamente todos los canales expresivos (manos, rostro, postura, espacio de señas), lo que explica su rezago frente al reconocimiento y la traducción.

El creciente uso de enfoques multimodales y de modelos generativos (GAN, difusión, LLM) confirma que la producción automática de lenguaje de señas avanza hacia mayor naturalidad gestual y coherencia temporal. Sin embargo, los modelos generativos analizados presentan una limitación común: ninguno modela simultáneamente todos los canales expresivos del lenguaje de señas (manos, rostro, postura corporal y espacio de señas), lo que reduce la fidelidad lingüística de las secuencias producidas. Esta limitación explica, en parte, por qué los sistemas de producción automática continúan rezagados respecto a los de reconocimiento y traducción.

En relación con la bidireccionalidad, solo 20 de los 50 estudios (40%) abordan explícitamente arquitecturas que integran reconocimiento, traducción y producción. Iniciativas como SignON y los talleres WMT-SLT constituyen avances relevantes, pero permanecen en fases experimentales debido a la escasez de datasets multilingües de gran escala y a la ausencia de métricas estandarizadas para evaluar la calidad semántica y gestual de las traducciones. Esta escasez de datos explica de forma directa por qué la mayoría de los sistemas revisados dependen de un número reducido de corpus (PHOENIX-2014T, How2Sign), lo que introduce un sesgo lingüístico que limita la generalización de los modelos a otras lenguas de señas.

La dependencia de pocos corpus afecta la validez de generalización de los sistemas. PHOENIX-2014T y How2Sign, los datasets predominantes, se limitan a la lengua de señas alemana y americana respectivamente, cada una con estructura gramatical y léxico propios; un modelo entrenado exclusivamente sobre ellos incorpora sesgos que no se replican en otras lenguas, como la colombiana o la mexicana, por lo que sus mejoras de precisión deben interpretarse como avances válidos en ese dominio, no como generalización transferible a otras comunidades sordas.

JWSign (Gueuwou et al., 2023), con más de 20 lenguas de señas, intenta abordar esta limitación, pero su heterogeneidad de calidad entre lenguas reproduce, a menor escala, el mismo problema. Mientras la investigación siga evaluando sus arquitecturas sobre uno o dos corpus de referencia, los avances reportados corren el riesgo de sobreestimar la madurez real del campo.

Más allá de la cobertura lingüística, los corpus predominantes se construyeron a partir de un número limitado de señantes en condiciones controladas de grabación, lo que introduce riesgo de sesgo algorítmico: los modelos tienden a optimizarse para ese perfil y esas condiciones, sin generalizar a señantes con estilos gestuales distintos o entornos menos controlados. Esto es coherente con la debilidad identificada en el análisis DOFA sobre la limitada validación en contextos reales.

Esta brecha entre condiciones de entrenamiento y de uso real constituye uno de los principales desafíos de implementación identificados. Un sistema con métricas elevadas sobre un dataset de referencia no garantiza desempeño equivalente frente a usuarios sordos reales, cuya variabilidad gestual excede la representada en los corpus disponibles. A esto se suma que la mayoría de los estudios no incorporan procesos de validación con señantes nativos, intérpretes o expertos lingüísticos, lo que exige diversificar los datos de entrenamiento e integrar a los usuarios finales como parte activa de la evaluación.

El aporte de esta revisión radica en vincular explícitamente las limitaciones arquitectónicas, la disponibilidad de datos y la validación en contextos reales como dimensiones interdependientes, lo que

explica por qué la comunicación bidireccional plenamente funcional continúa siendo una meta pendiente pese a la madurez de los sistemas unidireccionales.

El análisis DOFA (Tabla 5) sintetiza estos hallazgos: la solidez metodológica del campo (uso de protocolos PRISMA/Kitchenham) y la actualidad de la revisión constituyen fortalezas, mientras que la escasez de datasets y la limitada validación en contextos reales representan las principales debilidades. El auge de los modelos fundacionales multimodales ofrece una oportunidad clara para avanzar hacia la traducción sin glosas y la estandarización de métricas, aunque la velocidad de avance de la IA generativa y la fragmentación lingüística entre las distintas lenguas de señas constituyen amenazas relevantes para la consolidación del campo.

Tabla 5

Análisis DOFA del estado del arte en traducción automática del lenguaje de señas

Fortalezas			Oportunidades
Rigor metodológico (PRISMA/Kitchenham)	que garantiza la replicabilidad.	consolidado hasta 2025, incluyendo modelos generativos y LLM.	Integración con LLM para profundizar en la traducción sin glosas. Oportunidad de proponer un marco de métricas unificado ante la falta de estándares.
Debilidades			Amenazas
Escasez de datasets multilingües a gran escala. Falta de validación en contextos sociales o digitales reales.			Obsolescencia rápida por la velocidad de avance de la IA generativa. Fragmentación lingüística entre las distintas lenguas de señas.

En conjunto, estos hallazgos permiten explicar por qué, pese al nivel de madurez alcanzado por los sistemas unidireccionales, el desarrollo de tecnologías verdaderamente bidireccionales continúa siendo un campo emergente con amplias oportunidades de investigación.

CONCLUSIONES

Esta revisión sistemática permite analizar el estado actual de la investigación sobre sistemas de inteligencia artificial aplicados al reconocimiento, traducción y producción automática del lenguaje de señas entre 2018 y 2025, a partir de 50 estudios seleccionados mediante un protocolo estructurado de búsqueda, filtrado y evaluación de calidad.

En relación con la primera pregunta de investigación, se identifica que las tecnologías predominantes corresponden a arquitecturas Transformer, seguidas de enfoques multimodales y modelos generativos avanzados, lo que evidencia una transición tecnológica sostenida hacia modelos capaces de integrar múltiples fuentes de información visual y lingüística.

Respecto a la segunda pregunta, las principales contribuciones se concentran en tres líneas: el reconocimiento automático mediante estimación de pose y redes de grafos, la traducción automática Sign→Text mediante Transformers, y la producción automática Text→Sign mediante modelos generativos. Estas contribuciones mejoran de forma consistente la precisión semántica y la naturalidad gestual de los sistemas desarrollados.

En cuanto a la tercera pregunta, persisten desafíos estructurales relacionados con la escasez de datasets multilingües de gran escala, la ausencia de métricas de evaluación estandarizadas para el lenguaje de señas y la limitada validación de los sistemas en contextos reales de comunicación. Estos desafíos explican por qué los sistemas bidireccionales, pese a su relevancia para la inclusión comunicativa plena, continúan siendo minoritarios frente a los enfoques unidireccionales.

Se concluye que la inteligencia artificial posee un potencial considerable para mejorar la accesibilidad comunicativa de las personas sordas, pero que su consolidación como herramienta de inclusión social depende del desarrollo de datasets más diversos, arquitecturas verdaderamente bidireccionales y metodologías de evaluación adaptadas a las particularidades lingüísticas y visuales del lenguaje de señas.

Como limitación del estudio, se reconoce que la estrategia de búsqueda, aunque abarcó bases de datos ampliamente reconocidas, pudo no recuperar estudios publicados en fuentes no indexadas o en literatura gris. Asimismo, la heterogeneidad metodológica entre los estudios analizados dificulta la comparación cuantitativa directa entre propuestas.

Para futuras investigaciones, se recomienda promover la creación de datasets multimodales y multilingües de mayor escala, fortalecer el desarrollo de arquitecturas bidireccionales que operen en tiempo real, diseñar métricas de evaluación específicas para el lenguaje de señas, e incorporar validaciones en escenarios reales de comunicación con la participación de personas sordas, intérpretes y expertos lingüísticos.

AGRADECIMIENTOS

A Dios, por concederme la vida, la salud, la fortaleza y las oportunidades necesarias para alcanzar cada una de mis metas. Su guía y bendiciones han sido el fundamento que me permitió culminar esta etapa de mi formación profesional.

A mis padres, Omairan del Socorro Ibarra Fernández y Hernán Claudio Obando Andrade, por su amor incondicional, sus enseñanzas, sus sacrificios y el apoyo constante que me han brindado a lo largo de mi vida. Gracias por creer siempre en mí y por inculcarme los valores que hoy orientan mi camino.

A mi hermano, John Lenin Obando Ibarra, por su respaldo permanente, su confianza y por motivarme a superar cada desafío. Su compañía y ejemplo han sido una fuente de inspiración para continuar creciendo personal y profesionalmente.

A mis familiares y amigos, por su apoyo, comprensión y palabras de aliento durante este proceso. Su compañía en los momentos de mayor exigencia fue fundamental para mantener la motivación y perseverar hasta alcanzar este logro.

Finalmente, expreso mi más sincero agradecimiento a todas las personas que, de una u otra manera, contribuyeron a la realización de este trabajo y a mi crecimiento académico y personal.

REFERENCIAS

- Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., ... Vogler, C. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 16–31. <https://doi.org/10.1145/3308561.3353774>
- Camgöz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7784–7793).
- Camgöz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Duarte, A., et al. (2021). How2Sign: A large-scale multimodal dataset for continuous American Sign Language. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

- Fang, Y., Li, X., Zhang, H., & Zhao, L. (2024). SignLLM: Large language models for multimodal sign language understanding and generation. arXiv. <https://doi.org/10.48550/arXiv.2401.02885>
- Gueuwou, S., Siake, S., Leong, C., & Müller, M. (2023). JWSign: A highly multilingual corpus of Bible translations for more diversity in sign language processing. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 9907–9927). Association for Computational Linguistics.
- Guo, D., Li, X., & Wang, H. (2024). Diffusion models for sign language video generation. arXiv. <https://doi.org/10.48550/arXiv.2402.01743>
- Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering (EBSE Technical Report EBSE-2007-01). Keele University and University of Durham.
- Müller, S., et al. (2022). Findings of the WMT 2022 shared task on sign language translation. Proceedings of the Conference on Machine Translation (WMT).
- Müller, S., et al. (2023). Findings of the WMT 2023 shared task on sign language translation. Proceedings of the Conference on Machine Translation (WMT).
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2018). Sign language recognition using convolutional neural networks. European Conference on Computer Vision Workshops (ECCV Workshops).
- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition using pose-based convolutional neural networks. *IEEE Transactions on Multimedia*.
- Saunders, B., Camgöz, N. C., & Bowden, R. (2020). Everybody sign now: Translating speech to photo-realistic sign language video. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Saunders, B., Camgöz, N. C., & Bowden, R. (2022). Signing at scale: Learning to co-articulate sign language production. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Stoll, S., Camgöz, 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*.
- Vandeghinste, V., et al. (2023). The SignON project: Bridging communication between deaf and hearing communities. Proceedings of the European Association for Machine Translation (EAMT).
- Wong, R., Zhao, Q., & Li, H. (2024). Sign2GPT: Leveraging large language models for sign language translation. arXiv preprint.
- Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. Proceedings of the AAAI Conference on Artificial Intelligence.
- Ye, X., Zhou, W., & Li, H. (2024). Reducing representation density in gloss-free sign language translation. *Advances in Neural Information Processing Systems (NeurIPS)*.
- Yin, K., Read, J., & Liu, H. (2021). Gloss-free sign language translation using deep neural networks. Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL).
- Zhang, Y., Zhao, Q., & Li, H. (2024). Bidirectional sign language translation framework based on multimodal transformers. arXiv preprint.
- Zhou, W., Li, H., & Wang, Y. (2021). Improving sign language translation with back-translation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).