

Lenguaje como sistema predictivo: hacia una perspectiva semiótico-informática del discurso

Language as a predictive system: Towards a semiotic-informatical perspective on discourse

MARGARITA FLOR DE MARÍA MÉNDEZ-OCHAITA*

Este artículo propone un marco teórico para explorar las condiciones contemporáneas de producción de conocimiento y las prácticas discursivas en entornos mediados por inteligencia artificial (IA), así como algunas de sus implicaciones. Sostiene que la mediación algorítmica introduce una lógica predictiva del lenguaje que reconfigura las formas de producción discursiva. A partir de ello, plantea las bases de una perspectiva semiótico-informática orientada al estudio del discurso mediado por IA. El problema central radica en la escasa articulación entre la concepción del lenguaje como sistema predictivo, desarrollada en ámbitos informáticos, y su comprensión como práctica social en los estudios del discurso. Desde un enfoque teórico y transdisciplinario, el estudio integra aportes de las literacidades socioculturales, los estudios del discurso, la lingüística computacional y la informática. Mediante un análisis conceptual, examina principios técnicos, estructurales y funcionales de los sistemas de generación textual basados en IA y los vincula con una lectura semiótica de la producción de significados. Como resultado, propone comprender la lógica predictiva como una nueva forma de mediación tecnológica y plantea un marco conceptual que articula la dimensión sociotécnica de los modelos generativos con los estudios del discurso y las literacidades digitales.

Palabras clave:

estudios del discurso, literacidades digitales, inteligencia artificial, sistemas sociotécnicos, procesamiento de lenguaje natural

Recibido: 24 de marzo de 2026 | **Aceptado para su publicación:** 17 de julio de 2026 |

Publicado: 31 de agosto de 2026

Cómo citar: Méndez-Ochaita, M. F. M. (2026). Lenguaje como sistema predictivo: hacia una perspectiva semiótico-informática del discurso. *Sinéctica, Revista Electrónica de Educación*, (67), e1856. <https://doi.org/10.31391/AXTE8467>

This paper proposes a theoretical framework for exploring contemporary conditions of knowledge production and discursive practices in environments mediated by artificial intelligence (AI), as well as some broader implications of its use. It argues that algorithmic mediation introduces a predictive logic of language that reconfigures forms of discursive production. On this basis, the paper outlines a semiotic-computational perspective for the study of AI-mediated discourse. The central problem lies in the limited articulation between conceptions of language as a predictive system in computational fields and as a social practice in discourse studies, which hinders critical examination of these technologies, particularly in academic contexts. Adopting a theoretical and trans-disciplinary approach, the study draws on sociocultural perspectives on literacies and discourse, together with contributions from computational linguistics and computer science. Through conceptual analysis, it examines the technical, structural, and functional principles of AI-based text generation systems and relates them to a semiotic understanding of meaning-making in digital environments. It also contrasts language as a sociotechnical system with language as situated practice. The paper ultimately proposes a conceptual framework that integrates the sociotechnical dimension of generative language models with discourse and literacy studies while highlighting implications for academic writing, authorship, and digital literacy practices.

Keywords:

discourse studies, digital literacies, artificial intelligence, sociotechnical systems, natural language processing

* Doctora en Ciencias en la Especialidad de Investigaciones Educativas por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional de México. Profesora e investigadora de la Facultad de Filosofía y Letras de la Universidad Autónoma de Tlaxcala. Miembro del Sistema Nacional de Investigadoras e Investigadores de México (SNII-SECIHTI). Líneas de investigación: literacidades académicas y digitales, formación de investigadores, trayectorias académicas y procesos de institucionalización científica en México. Correo electrónico: flor.mendez@uatx.mx/
<https://orcid.org/0000-0001-7601-4570>



INTRODUCCIÓN

En la actualidad, los avances en informática, en particular en el desarrollo de sistemas de procesamiento automático del lenguaje natural (PLN) han representado una transformación de la economía mundial con un impacto visible en diversos ámbitos de la vida social, entre ellos, la industria, la banca, la salud, el comercio, la comunicación y la educación (Kaur et al., 2022; Sanabria-Navarro et al., 2023). En este amplio contexto, la inteligencia artificial (IA) ha consolidado un modelo operativo según el cual comprender y producir lenguaje implica anticipar probabilísticamente unidades lingüísticas a partir de patrones aprendidos en grandes corpus de datos. Esta modelización algorítmica del lenguaje, y de una altísima significación estadística, no solo ha impulsado el desarrollo de las tecnologías digitales, sino también está reformulando los marcos conceptuales desde los cuales se piensa el lenguaje. En este sentido, se parte de la hipótesis de que las transformaciones asociadas a la IA representan innovaciones técnicas que reconfiguran las formas en que los discursos se producen, legitiman y circulan en entornos digitales e implican cambios significativos en la educación superior vinculados con intereses socioeconómicos más amplios (Bearman et al., 2023).

En el ámbito ingenieril informático, la predicción del lenguaje se ha consolidado mediante el aprendizaje de máquinas o automático (*machine learning*, ML) (Forero-Corba y Bennasar, 2024). En particular, la lingüística computacional, rama interdisciplinaria orientada a la elaboración de modelos computacionales capaces de reproducir uno o más aspectos del lenguaje humano (Domínguez, 2002), ha contribuido mediante la tecnología de PLN (Schaub, 2020).

Por su parte, tradiciones de la lingüística y de los estudios del discurso han ayudado a comprender estructuras, funciones y usos del lenguaje. Entre las corrientes principales que conciben el lenguaje como sistema, están la racionalista-estructuralista, que lo entiende como sistema autónomo de reglas formales y competencia lingüística del hablante normal de una lengua (Chomsky, 1969/1991), y la corriente sistémico-funcional desarrollada desde la semiótica social y la lingüística funcional, que concibe un sistema de opciones mediante el cual los hablantes construyen significados según sus condiciones sociales e históricas (Halliday, 1982/2017), además de enfoques socioculturales sobre el uso de lenguaje en contextos específicos que conciben la lectura y la escritura como prácticas discursivas y situadas (Street, 2009).

Sin embargo, existe una divergencia significativa entre las concepciones del lenguaje de campos orientados al desarrollo tecnológico y aquellos dedicados al análisis del discurso. Mientras que en ámbitos informáticos el lenguaje natural es entendido como producto de un sistema algorítmico de predicción semiótico-probabilístico, en los estudios del discurso continúa analizándose desde marcos conceptuales que priorizan las dimensiones sociales, culturales y contextuales de la producción de significado, pero que todavía no incorporan de manera plena las dimensiones estructurales de dichas tecnologías. Esta falta de integración resulta problemática porque genera una asimetría epistemológica entre ambos ámbitos: por un lado, prevalece la modelización avanzada del lenguaje y, por otro, el estudio de las prácticas discursivas mediadas por estas tecnologías sin marcos conceptuales que integren teórica y formalmente su dimensión técnica.

En el contexto actual, esta divergencia se vuelve relevante, ya que la lógica predictiva del lenguaje que estructura los sistemas digitales resulta desconocida para la mayoría de los usuarios que interactúan cotidianamente con estas tecnologías. Por ello, su estudio es relevante en ámbitos académicos de producción, circulación y validación del conocimiento. Al usar estas tecnologías sin una comprensión crítica, se genera una brecha entre quienes diseñan los sistemas y quienes interactúan con ellos, lo que afecta la agencia discursiva y los modos actuales de construcción socio-cultural del significado (Fairclough, 2003; Van Dijk, 2019).

En este contexto, en este artículo analizamos de manera teórica el funcionamiento de los modelos predictivos de lenguaje partiendo de un enfoque sociocultural de las literacidades académicas desarrollado originalmente en estudios de corte etnográfico por autores como Street (2004) y Zavala (2009), y que destaca su comprensión como prácticas sociales. El objetivo es proponer un marco teórico para examinar las condiciones de producción y mediación tecnológica del discurso con IA, reconfiguradas por la lógica predictiva del lenguaje, así como algunas de sus implicaciones generales de uso. Para ello, examinamos y describimos los componentes estructurales del entorno digital basados en IA que condicionan las prácticas letradas contemporáneas y la construcción de significados discursivos. Estas tecnologías resultan problemáticas, ya que, por un lado, amplían el acceso y la participación en prácticas discursivas de los usuarios; y por otro, aumentan la dependencia tecnológica, al disminuir la agencia discursiva de los usuarios (Zavala, 2011).

La tesis central del artículo sostiene como hipótesis que el carácter predictivo del lenguaje en los sistemas de IA constituye una nueva fase en la mediación tecnológica que reconfigura las prácticas discursivas y los procesos de significación en los entornos digitales. Sin embargo, la falta de articulación entre la concepción del lenguaje como sistema predictivo desarrollada en el ámbito informático y su abordaje como práctica social en los estudios del discurso, dificulta la comprensión de los efectos generales de estas tecnologías en las prácticas discursivas y en la producción de conocimiento, lo cual es relevante para el ámbito académico en estudios posteriores. Frente a esta situación, el artículo propone avanzar hacia una nueva perspectiva semiótico-informática del lenguaje que permita integrar teóricamente las dimensiones técnicas y socioculturales sin subordinar una a la otra, con el fin de construir un marco analítico más adecuado para examinar en futuras investigaciones las transformaciones actuales de las prácticas discursivas digitales y sus nuevas formas de significación.

Para abordar este problema, el artículo se organiza en cuatro secciones. La primera presenta el enfoque teórico y metodológico que sustenta la investigación. La segunda describe en forma breve la evolución del campo de la inteligencia artificial y contrasta dos concepciones sobre el lenguaje, una de carácter técnico desarrollada en el ámbito de los sistemas computacionales, y otra de carácter situado en la interacción cotidiana entre usuarios y tecnologías digitales generadoras de textos. A partir de ejemplos de tareas de predicción integradas en entornos digitales cotidianos, se evidencian efectos para la agencia discursiva de los sujetos y la construcción de significados del lenguaje en diversos ámbitos.

La tercera sección describe la arquitectura básica de los sistemas de procesamiento de lenguaje natural basados en IA, sus principales componentes estructurales y los modelos basados en grandes volúmenes de datos, con la finalidad de relacionar

teóricamente el funcionamiento sociotécnico de los modelos predictivos con la noción de prácticas discursivas digitales. La cuarta y última sección aporta una aproximación teórica inicial que caracteriza el lenguaje mediado por IA como un sistema predictivo y traza las bases de una perspectiva semiótico-informática del discurso. Por último, el artículo cierra con las conclusiones del trabajo.

PERSPECTIVA TEÓRICA QUE SUSTENTA LA INVESTIGACIÓN

El estudio de enfoque teórico-conceptual y carácter transdisciplinar parte de los marcos de los estudios del discurso y la perspectiva sociocultural de las literacidades, y se desarrolla con aportes de la lingüística computacional y la informática para comprender el funcionamiento de los sistemas de IA y las formas de mediación algorítmica a fin de construir un marco conceptual compartido que integre, y trascienda, los aportes particulares de las disciplinas involucradas (Henao-Villa et al., 2017).

El estudio concibe las literacidades como prácticas discursivas situadas en contextos específicos de uso, históricamente configuradas por ideologías, estructuras normativas, relaciones de poder y tecnologías (Street, 2004, 2009; Zavala, 2009, 2011). Por lo mismo, entiende la producción de significados como resultado de prácticas sociales vinculadas a culturas y condiciones históricas y materiales particulares.

Este planteamiento dialoga con los estudios de literacidad digital que permiten comprender las literacidades digitales como prácticas de comunicación, relación, pensamiento y construcción identitaria vinculadas al uso de medios digitales (Jones y Hafner, 2021, 2012; Méndez-Ochaita, 2022; Méndez-Ochaita y Hernández-Ramírez, 2026). En este sentido, la noción de práctica no se limita al manejo instrumental de plataformas o dispositivos inteligentes, sino que implica la capacidad de interpretar, adaptar y resignificar las posibilidades y limitaciones que estas tecnologías introducen en la interacción social y en la producción de significados. De este modo, el estudio también incorpora una lectura semiótica social del lenguaje mediado por IA que permite entenderlo como parte de un sistema de significación relacionado con procesos lingüísticos por los cuales se transforma, se limita y se modifica la realidad social, y con dinámicas de persistencia y cambio social entre las cuales se construyen los significados (Halliday, 1982/2017).

Por otra parte, para examinar conceptualmente la relación entre los sistemas de generación del lenguaje basados en IA y las prácticas discursivas que se configuran en los entornos digitales, se asume que los procesos de producción, circulación e interpretación de los discursos no pueden comprenderse al margen de las condiciones sociotécnicas que los hacen posibles (Sánchez, 2021). Por ello, se retoma la noción de sistemas sociotécnicos de IA que permite comprender el lenguaje mediado por IA como resultado de la interacción entre usuarios, algoritmos, marcos institucionales y prácticas sociodiscursivas. Esta conceptualización destaca una agencia distribuida entre los usuarios que interactúan con infraestructuras tecnológicas y las organizaciones que regulan, diseñan y controlan estos sistemas, desarrollando algoritmos y reglas técnicas que configuran y adaptan su comportamiento (Kudina y Van de Poel, 2024).

En este marco, la relación entre tecnologías y prácticas discursivas se desarrolla a partir del concepto de mediación, entendido como el conjunto de procesos materiales y simbólicos que intervienen en la acción humana, la construcción de significados y

las formas de interacción social. En particular, la mediación algorítmica puede definirse como el conjunto de procesos mediante los cuales los algoritmos intervienen en la generación, selección, organización y circulación de contenidos, condicionando las formas en que estos son producidos, interpretados y compartidos. Desde esta perspectiva, la IA puede entenderse como un “agente cognitivo y simbólico que redefine la producción de sentido” en culturas digitales diversas (Estrada et al., 2026, p. 13).

EVOLUCIÓN DE LA INTELIGENCIA ARTIFICIAL COMO MEDIACIÓN SEMIÓTICA DEL LENGUAJE

La IA surge como una disciplina científica y tecnológica orientada a la creación de sistemas capaces de llevar a cabo tareas que requieren capacidades humanas. En este sentido, la noción de IA parte de la idea de construir máquinas “inteligentes” que funcionen de manera apropiada en su entorno (McCarthy et al., 2006, p. 11), lo que requiere la capacidad de interactuar mediante el lenguaje. Este enfoque evolucionó con aportes de informática, psicología, biología, neurología y filosofía hacia una perspectiva sistémica orientada al diseño de agentes computacionales que “actúan inteligentemente” en un entorno para lograr determinados objetivos (Poole et al., 1998, p. 1).

En forma posterior, se consolidó una perspectiva cognitiva basada en sistemas que “aprenden” de datos para realizar tareas que simulan el pensamiento humano, como razonamiento, percepción y comprensión de lenguaje natural (Forero-Corba y Bennasar, 2024, p. 21), lo cual marca un avance significativo para el desarrollo de estas tecnologías capaces de participar en prácticas comunicativas mediadas digitalmente.

Entre las principales funciones de estos sistemas se encuentra la toma de decisiones, la resolución de problemas y el aprendizaje automático (*machine learning*, ML) (Sanabria-Navarro et al., 2023). Su característica principal es su capacidad para predecir respuestas de forma automática y mejorar su rendimiento, sin necesidad de tener una programación explícita para cada regla (Forero-Corba y Bennasar, 2024; Soto et al., 2024).

En general, el desarrollo tecnológico de los sistemas de IA ha sido impulsado por tres avances fundamentales en sus componentes operativos: el desarrollo de algoritmos avanzados, el aumento de la capacidad de cómputo y el acceso a grandes volúmenes de datos (Franganillo, 2023). Como resultado de esta expansión, se ha consolidado una concepción sistémica cognitiva de la IA y los modelos de lenguaje en distintos campos de aplicación, entre ellos las ingenierías de sistemas informáticos y computacionales, la ciencia de datos y la minería de datos (véase figura 1).

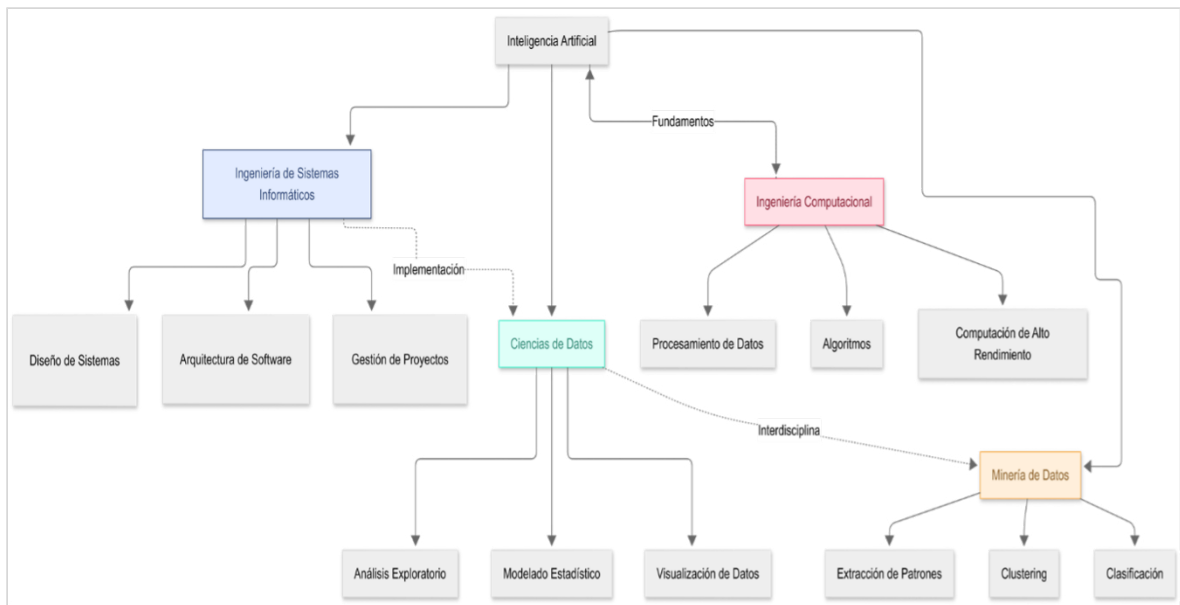


Figura 1. Campos de la inteligencia artificial. Fuente: Elaboración propia con *Mermaid*, herramienta basada en programación JavaScript a partir de texto plano y sintaxis de símbolos para diseño del formato.

En estos campos, el ML constituye una rama dedicada al desarrollo de algoritmos capaces de utilizar datos para detectar patrones y resumir determinados fenómenos. En particular, estos sistemas se entrenan para crear modelos predictivos mediante técnicas que se clasifican según el tipo procesamiento que se aplique a los datos para derivar patrones y tendencias de estos. Entre las técnicas de ML más usadas están la clasificación supervisada con datos etiquetados por los desarrolladores (Segura et al., 2022), el agrupamiento no supervisado que opera con datos no etiquetados (Taha et al., 2018), el aprendizaje por refuerzo que usa datos a partir de las interacciones con el entorno (Dietz et al., 2022) y el aprendizaje semisupervisado que combina datos etiquetados y no etiquetados (Forero-Corba y Bennasar, 2024; Salas-Pilco y Yang, 2022).

Igual que las anteriores, el aprendizaje profundo (*deep learning*, DL) es una técnica potente y de gran flexibilidad que permite a los sistemas mejorar con base en experiencia y datos. Se trata de una modalidad del ML que emplea redes neuronales artificiales organizadas en múltiples capas de nodos conectados entre sí, con el propósito de imitar el funcionamiento del cerebro humano. Una vez entrenados, estos algoritmos pueden realizar predicciones automáticas a partir de grandes volúmenes de datos (Salas-Pilco y Yang, 2022). Esta técnica funciona como una jerarquía anidada de conceptos y representaciones, donde cada concepto se define en relación con conceptos más simples y las representaciones más abstractas se calculan en función de las menos abstractas (Goodfellow et al., 2017).

En la actualidad, las aplicaciones de IA en educación abarcan procesos de evaluación, predicción, asistencia automatizada, tutoría y aprendizaje personalizado (Crompton y Burke, 2023; Sanabria-Navarro et al., 2023), y se sustentan en modelos de procesamiento de lenguaje que median la interacción entre usuarios y sistemas digitales.

Concepción técnica del lenguaje predictivo

El procesamiento del lenguaje constituye una rama de la IA orientada a producir lenguaje natural, anticipando unidades lingüísticas a partir de probabilidades y patrones previamente aprendidos en grandes corpus de datos. La predicción se basa en el análisis de estructuras lingüísticas con algoritmos avanzados que permiten interpretar el significado de las palabras. Entre sus aplicaciones más conocidas se encuentran los *chatbots* que procesan solicitudes en lenguaje cotidiano y ofrecen respuestas automáticas, así como los sistemas de traducción automática de idiomas (Salas-Pilco y Yang, 2022).

En términos operativos, el lenguaje es procesado mediante sistemas computacionales que articulan modelos matemáticos, cálculos estadísticos y optimización de datos, así como arquitecturas algorítmicas, DL y ML. El lenguaje conceptualizado como un sistema formal de estructuras o como una actividad situada, al ser modelado estadísticamente en el marco de la informática, se define como un sistema predictivo. A partir de ello, la producción discursiva se convierte en el resultado de una secuencia de probabilidades derivadas de patrones, regularidades y clasificaciones extraídas de datos masivos.

En tal sentido, estas mediaciones tecnológicas hacen posible que el lenguaje pueda ser anticipado y ajustado mediante algoritmos (Soto et al., 2024), por lo cual este pasa de ser considerado una competencia exclusivamente humana a una capacidad emergente del sistema de tipo “cognitivo”, es decir, una “inteligencia artificial”. De ahí la necesidad de reconfigurar los marcos teóricos que faciliten a los usuarios, no solo desarrollar sus capacidades para adaptarse a los entornos digitales, sino también para comprender un lenguaje digital universalizado sustentado en programas de IA (Ocaña-Fernández et al., 2019).

Concepción pragmática del lenguaje mediado por IA

Desde el enfoque sociocultural adoptado en este estudio, los usos del lenguaje constituyen prácticas sociales (Street, 2009; Zavala, 2009). En este sentido, la experiencia situada de los usuarios en entornos digitales mediados por IA se configura en interacciones concretas con procesos de negociación de significados que ocurren en escenarios sociales específicos, en los cuales los avances informáticos han contribuido a incorporar el lenguaje natural en tareas de escritura para mejorar una redacción, facilitar la comprensión de un texto extenso o traducirlo, entre otras (Beltrán y Rodríguez, 2020).

No obstante, la noción del lenguaje como sistema predictivo desarrollada desde ámbitos informáticos y computacionales no ha sido incorporada en forma plena en los marcos epistemológicos de los estudios del discurso. Por lo mismo, la predicción es más vista como una funcionalidad operativa del sistema, pero no tanto como una dimensión analítica para interpretar cómo los usuarios viven, interpretan, malinterpretan o resignifican estas mediaciones algorítmicas en su experiencia cotidiana. Las prácticas discursivas, aunque sean resultado de aceptar, modificar o resistir una sugerencia automática del sistema, no pueden reducirse a cálculos probabilísticos.

Esta asimetría de tipo conceptual es relevante en un entorno digital en el que las herramientas predictivas intervienen en las prácticas de lectura, escritura y producción de conocimiento. La divergencia entre los campos del desarrollo tecnológico y los dedicados al análisis del discurso configura lo que aquí hemos denominado una asimetría epistemológica entre concepciones de lenguaje mediado por IA. Mientras que, en el plano ingenieril y de la lingüística computacional, el lenguaje natural proviene de sistemas predictivos de modelación algorítmica y estadística, en el plano formativo y discursivo continúa siendo comprendido desde marcos tradicionales que destacan su dimensión social, pragmática e histórica, sin integrar el funcionamiento de las infraestructuras informáticas que median estas prácticas y los significados. En el siguiente apartado compartimos ejemplos que evidencian la función predictiva y asistencial del lenguaje subyacente en medios digitales.

Entornos digitales y ejemplos de predicción textual

La generación del lenguaje por sistemas de IA no permanece en el plano abstracto del modelado, sino que se materializa en dispositivos y aplicaciones de uso cotidiano que operan predictivamente en tiempo real, por ejemplo, en sistemas de autocompletado, autocorrección, generación automática, sugerencias de búsqueda y recomendaciones de contenido. El principio subyacente es similar: a partir de datos de diversas fuentes, el sistema calcula la opción más probable, ya sea la siguiente letra, palabra, frase, video o corrección, y la ofrece al usuario de forma anticipada a su acción.

El autocompletado, por ejemplo, sugiere palabras o segmentos antes de que el usuario concluya su enunciado. Otros sistemas, que median de forma extendida las prácticas de lectura y escritura en entornos digitalizados, también operan bajo una lógica semejante de optimización probabilística, y su función principal es ofrecer asistencia al usuario, como la autocorrección textual que reemplaza formas consideradas desviadas respecto de un patrón estadístico dominante, la generación automática de texto, la cual produce secuencias completas que demuestran coherencia discursiva, o los sistemas de recomendación que funcionan priorizando determinados textos, productos o trayectorias de lectura.

No obstante, estos procesos generativos no son autónomos, sino automatizados. La generación textual está condicionada por el entrenamiento de un modelo a partir de datos de entrada, y su capacidad de generar contenido se fundamenta en el aprendizaje de patrones presentes en dichos datos, así como en la posterior producción de resultados que reproducen esos patrones de formas nuevas guiadas por las instrucciones del usuario (García-Peñalvo y Vásquez-Ingelmo, 2023; García-Peñalvo et al., 2024). Debido a esta dependencia, cuando las interacciones se apoyan de modo continuo en datos previos, se refuerzan patrones que tienden a producir una homogenización discursiva y a introducir mayores sesgos.

Aunado a lo anterior, una problemática relevante para los estudios de discurso y literacidades es que la generación automática de contenido afecta en dos sentidos a las prácticas discursivas: por un lado, posibilita un amplio acceso y, por otro, puede afectar las capacidades de agencia de los usuarios al incrementar su dependencia a estas tecnologías. Estos efectos se corresponden también

con dos posturas discursivas predominantes en la educación superior frente a estas tecnologías: aquellas que las conciben como un imperativo social al que la universidad debe responder para un mayor beneficio formativo y aquellas que las interpretan como una alteración de la capacidad de agencia (Bearman et al., 2023).

En efecto, las herramientas predictivas facilitan la escritura a usuarios con distintos niveles de competencia lingüística y pueden reducir barreras de entrada a géneros académicos o profesionales, acelerando también los procesos de producción textual. Asimismo, la corrección automática puede asistir en la adecuación discursiva de textos académicos o profesionales, mientras que el autocompletado disminuye el esfuerzo de redacción en prácticas cotidianas con dispositivos móviles. No obstante, también pueden incrementar la dependencia de los usuarios y, con ello, limitar su capacidad de agencia. Al respecto, un estudio experimental reportó un fenómeno denominado “deuda cognitiva” asociado al uso de IA en la escritura de ensayos. Si bien los participantes obtuvieron puntuaciones altas en la prueba escrita, evidenciaron un esfuerzo neuronal menos coordinado y un sesgo hacia el léxico del sistema, comparado con quienes usaron la IA solo como apoyo para reescribir un texto previamente redactado sin este recurso (Kosmyna et al., 2025).

En este sentido, las herramientas predictivas tienden a desplazar parte del trabajo cognitivo de los usuarios hacia el sistema aumentando su dependencia. El usuario tiende a adoptar las sugerencias que se le presentan sin cuestionarlas, naturalizando la probabilidad estadística como criterio válido de pertinencia discursiva, el cual también es favorecido por altos niveles de coherencia textual. Como resultado, la ampliación del acceso puede derivar en una estandarización de las formas de expresión, reduciendo la diversidad de estilos y propiciando una progresiva delegación de las elecciones lingüísticas en las arquitecturas algorítmicas cuya lógica predictiva, criterios de operación y funcionamiento permanecen opacos para la mayoría de los usuarios.

Por tanto, la interacción entre el usuario y los sistemas de procesamiento de lenguaje basados en IA resulta clave para los procesos formativos, en la medida en que no solo median la producción textual, sino que también configuran las expectativas discursivas y estabilizan ciertas regularidades semióticas. Más que herramientas neutrales de asistencia, estos sistemas no solo influyen en las capacidades cognitivas, sino intervienen en lo que se considera pertinente, incluso legítimo, a nivel de una producción sociodiscursiva estabilizada, cuestión que abordamos en la siguiente sección.

PERSPECTIVA SEMIÓTICO-INFORMÁTICA DEL DISCURSO MEDIADO POR IA

Conceptos y fundamentos

Esta propuesta busca relacionar teóricamente las prácticas sociales del lenguaje con los fundamentos técnicos del procesamiento de lenguaje natural (PLN) basados en IA y modelos de aprendizaje (*large language model* [LLM, modelo de lenguaje de gran escala] y DL) en un marco integrado que permita interpretar la producción discursiva en los sistemas de IA, considerando sus componentes estructurales: informático, computacional-probabilístico y contextual, que revisaremos con más detalle

a continuación, y las mediaciones a través de las cuales se organizan, transforman y producen sentidos discursivos con base en datos ponderados por modelos algorítmicos e interfaces.

A diferencia de otros enfoques centrados en los componentes operacionales del PLN o de análisis lingüísticos tradicionales, esta propuesta aborda el lenguaje como un sistema de producción semiótico-informática en el cual las mediaciones algorítmicas, en la selección de datos de entrenamiento, la intervención humana o el diseño de las interfaces, entre otras, condicionan las formas de significación discursiva y de agencia de los usuarios.

A nivel operativo, la arquitectura de estos sistemas materializa soluciones informáticas complejas que integran datos, algoritmos y modelos de aprendizaje cuya finalidad es simular capacidades cognitivas humanas como el razonamiento, el aprendizaje, la percepción y la toma de decisiones, y se estructuran en tres componentes básicos: la infraestructura, que incluye servidores, redes, almacenamiento y virtualización de cualquiera de estos componentes; el sistema de procesamiento de datos que incluye algoritmos y modelos predictivos e inferenciales; y una interfaz que se despliega para diferentes capacidades del usuario por medio de aplicaciones generales, *chatbots*, asistentes y espacios diseñados para la programación de aplicaciones (*application program interface*, API) (Porrás et al., 2018).

Estos sistemas ejecutan tareas específicas como análisis predictivo, valoración, clasificación, adaptación (ML), personalización (elaboración de perfiles), asistencia (reconocimiento de voz, *chatbots*) o reconocimiento de imágenes (visión artificial) (Crompton y Burke, 2023; Salas-Pilco y Yang, 2022; Zawacki-Richter et al., 2019). Para ello, procesan datos ingresados por los usuarios en la interacción o extraídos de diversas fuentes, entre los que se incluyen datos de uso y contexto del usuario, así como datos masivos. En el ámbito del PLN, mediante una serie de transformaciones, el modelo predictivo convierte al lenguaje humano en estructuras matemáticas, las procesa y luego reconvierte esas estructuras en lenguaje natural coherente (Beltrán y Rodríguez, 2020).

De este modo, el PLN puede entenderse como el campo de integración de las técnicas de IA con el fin de interpretar, manipular y comprender el lenguaje humano (Giraldo y Orozco, 2023), en el que convergen mediaciones algorítmicas que transforman el lenguaje en estructuras computacionales y luego las reconfiguran en discurso. Analizar su estructura y funcionamiento con una lectura semiótica social permite, por tanto, comprender los mecanismos a través de los cuales se producen y estabilizan los significados en los sistemas generativos de texto basados en IA.

Por último, la propuesta incorpora un componente crítico vinculado a la infraestructura tecnológica y las relaciones de poder de las empresas que controlan servicios de IA, lo que posibilita, por un lado, conceptualizar los discursos producidos con IA como lenguaje predictivo que reproduce regularidades culturales y jerarquiza significados, y por otro, problematizar implicancias generales de la personalización algorítmica en la producción y circulación del conocimiento.

Construcción socioinformática del significado mediante procesamiento de lenguaje natural-PLN

En la última década, el ámbito del PLN ha experimentado una transformación significativa, impulsada por avances en los modelos de lenguajes, cambios en las infraestructuras computacionales, un creciente interés en el ML y la necesidad de generar lenguaje natural de manera más efectiva (Delso et al., 2025; Young et al., 2018).

La figura 2 muestra el flujo de datos en una arquitectura de PLN. El diagrama incluye los tres componentes básicos del sistema: interfaz de usuario, proceso algorítmico con modelos de IA e infraestructura de almacenamiento. Asimismo, representa las operaciones generales que intervienen en las distintas fases de la producción discursiva mediada por IA, lo que permite comprender, a su vez, los principios del PLN que sustentan el funcionamiento de la IA generativa, los cuales describimos a continuación.

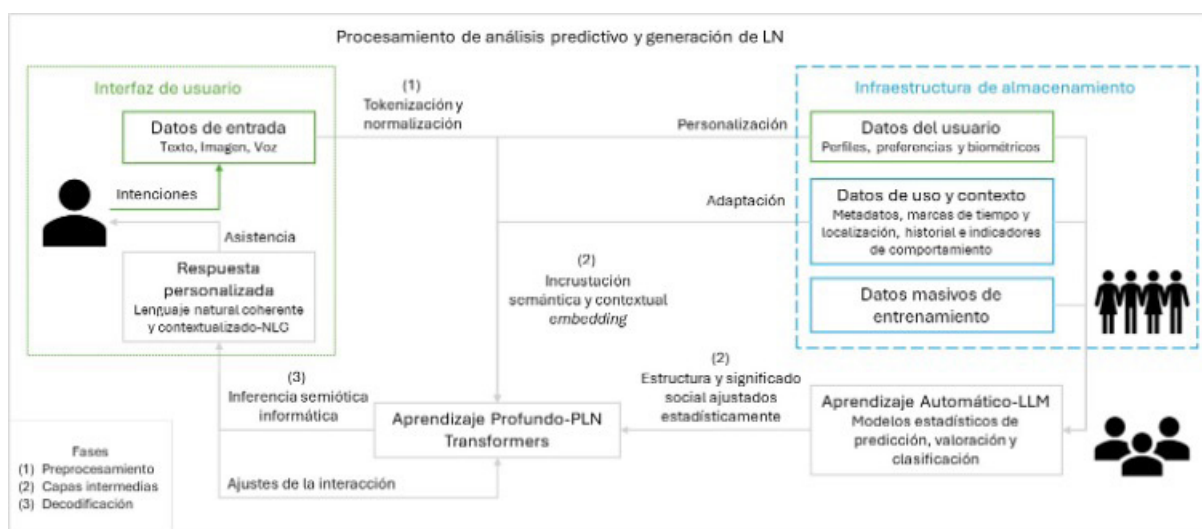


Figura 2. Arquitectura de PLN como construcción semiótico-informática del discurso mediado por IA.

El proceso básico de PLN inicia con la adquisición de datos. Estos pueden ser textos digitales provenientes de diferentes bases de datos que incluyen documentos, libros, foros o redes sociales, conversiones de audio a texto; datos multimodales derivados de la interacción con el usuario; y registros de uso y contexto almacenados en infraestructuras propias o virtualizadas. La selección del corpus determina qué voces, registros y perspectivas ingresan al sistema, de modo que la realidad aprendida por el modelo siempre es una construcción estadística basada tanto en datos masivos como personales que se genera mediante un proceso de al menos tres fases relacionadas con el DL: preprocesamiento textual, capas intermedias de codificación contextual y modelado semiótico, así como decodificación, las cuales se analizan con mayor profundidad a continuación.

La primera fase de preprocesamiento consiste en la descomposición del texto de entrada en unidades procesables. Esto incluye operaciones de limpieza, normalización y tokenización que implican decisiones normativas sobre lo que se considera ruido en el discurso, por ejemplo: palabras con escaso contenido semántico, errores

ortotipográficos o repeticiones. Además, el algoritmo determina las variables que se estandarizan y los elementos que se excluyen. Como resultado, convierte el texto de entrada en una secuencia de datos estructurada para su procesamiento computacional. La representación vectorial del texto se lleva a cabo mediante incrustaciones (*embedding*) lingüísticas. Este proceso es conocido como tokenización y transforma palabras y secuencias textuales en vectores numéricos que capturan similitudes semánticas y contextuales de unidades lingüísticas (Atkinson-Abutridy, 2023a; Porras et al., 2018).

Sobre esta base, la fase intermedia de codificación contextual profundiza esta lógica al modelar dependencias sintácticas y semánticas en secuencias completas, capturando coherencia discursiva y variaciones de sentido según el entorno lingüístico. De ese modo, el significado global del discurso se configura como una construcción dinámica y contextualizada (Atkinson-Abutridy, 2023a).

En arquitecturas modernas de DL basadas en el modelo Transformer, dichas representaciones incorporan dependencias de largo alcance que amplían estas capacidades, lo que permite modelar relaciones complejas dentro de una secuencia (Fúquene, 2024; Wolf et al., 2020). En específico, los modelos Generative Pre-trained Transformer se caracterizan por tener estructuras compuestas por múltiples capas con cientos de millones de parámetros que deben ajustarse durante el entrenamiento y requieren corpus de gran escala, integrados por miles de millones de palabras. Dado que estas capacidades de generación textual son comparables a las humanas, su impacto cultural, educativo y social ha ocasionado amplios debates. En esta fase de modelización el significado se construye como una relación probabilística, ya que una palabra adquiere valor por su posición en un espacio vectorial y por su proximidad estadística a otras (Giraldo y Orozco, 2023).

En continuidad con lo anterior, el modelado neuronal mediante DL, entrenado con grandes corpus de datos, permite al sistema identificar regularidades en datos etiquetados, detectar agrupamientos y anomalías en datos no etiquetados, y ajustar parámetros mediante retropropagación del error (Forero-Corba y Bannasar, 2024). En esta etapa, la predicción se efectúa como un ajuste iterativo que optimiza los resultados respecto a una salida proyectada. Aunque se trata de operaciones estadísticas, estas reproducen distintos niveles de análisis lingüístico: morfológico y léxico mediante la identificación de unidades y variaciones; sintáctico en las relaciones estructurales, semántico en la construcción de significado contextual, identificación de roles y relaciones y pragmático en la integración del discurso adecuado al contexto comunicativo y la intención del usuario. Así, aunque el sistema no se ejecuta por capas explícitas, su funcionamiento en el espacio probabilístico las realiza.

Finalmente, la tercera fase de decodificación crea texto coherente a partir de capas de salida que seleccionan y calculan la probabilidad de cada token hasta completar una respuesta. Este proceso, conocido como generación de lenguaje natural (*natural language generation*), produce secuencias que mantienen una correlación estadística con el contexto previo. El sistema produce regularidades probabilísticas condicionadas por el entrenamiento y el contexto de uso. Sin embargo, este proceso no implica una comprensión por parte del sistema en el sentido estricto de las capacidades humanas (Atkinson-Abutridy, 2023a, 2023b), sino valor asistencial para el usuario. La mediación algorítmica en la construcción del lenguaje dependerá de

estimaciones probabilísticas a partir de los datos que configuran el espacio discursivo, el entrenamiento de los modelos que instituye normas implícitas, la adaptación al usuario o personalización del discurso, y la interfaz que presenta el discurso final. Así, la predicción del lenguaje se convierte en un texto visible que no solo es cálculo estadístico o simulación, sino producción discursiva informática.

En síntesis, el fundamento general del PLN consiste en modelar procesos cognoscitivos que intervienen en la comprensión del lenguaje, con el objetivo de diseñar sistemas que realicen tareas lingüísticas complejas como traducción, resumen o recuperación de información. No obstante, estas capacidades en realidad son aproximaciones a distintos niveles de análisis morfológico, sintáctico, semántico y pragmático (Cortez et al., 2009) que dependen, en primer lugar, de modelos de lenguaje contruidos con grandes corpus, aspecto que se desarrolla en el siguiente apartado.

Lenguaje como modelo basado en datos masivos (large language models, LLM)

Los LLM representan una evolución del PLN basada en arquitecturas neuronales profundas entrenadas con volúmenes masivos de datos, y capaces de operar con grandes cantidades de parámetros. Su funcionamiento inicia con la recolección de datos relevantes, corpus de dominio público y textos digitales diversos, y sigue con procesos de limpieza y detección de anomalías que eliminan duplicados o filtran contenido no deseado.

Posteriormente, la etapa de entrenamiento a gran escala implementa una fase de modelado estadístico por medio de diferentes técnicas de aprendizaje: supervisado y no supervisado, para identificar patrones, tendencias y repeticiones en datos etiquetados, así como similitudes y agrupamientos en datos no etiquetados (Atkinson-Abutridy, 2023b).

Sobre esta base, la fase de interacción por refuerzo introduce retroalimentación humana en forma de recompensas y penalizaciones (Forero-Corba y Bennasar, 2024), alineando el comportamiento del modelo con expectativas sociales y normativas.

Por último, la etapa de evaluación del modelo valida su desempeño en tareas específicas con métricas de precisión, error y ajuste paramétrico. Para su despliegue, el modelo requiere interfaces conversacionales, API y servicios en la nube. En conjunto, estas etapas y fases cumplen funciones algorítmicas operativas, pero pueden ser vistas como instancia de mediación semiótico-probabilística a través de la cual se decide qué datos cuentan, qué patrones se consolidan y qué comportamientos se privilegian.

Por lo anterior, un LLM puede entenderse como un paquete estadístico que funciona como mecanismo de reproducción y reconfiguración cultural, emulando conocimiento a partir de distribuciones de probabilidad sobre secuencias lingüísticas. Así, cuando el usuario interactúa con un modelo de PLN, el sistema recibe un *prompt* que funciona como acto discursivo inicial, lo codifica en representaciones vectoriales, calcula distribuciones probabilísticas y genera texto token a token. El resultado es la materialización computacional de expectativas lingüísticas, ya que el modelo selecciona una trayectoria discursiva entre múltiples posibles, reajustando sus parámetros de entrenamiento con datos masivos.

En este sentido, la predicción en el sistema de IA constituye una intervención socioinformática en la producción de sentido. Las respuestas reproducen y, en algunos casos, reconfiguran patrones culturales dominantes presentes en el corpus de entrenamiento (Salas-Pilco y Yang, 2022). De este modo, el sistema convierte datos masivos en discurso situado, operando como dispositivo que traduce estructuras estadísticas en enunciación. La arquitectura técnica de los PLN y los LLM funciona como una cadena de transformaciones matemáticas basada en datos de lenguaje natural que configura discursos, define normas implícitas y materializa la predicción como texto operativo caracterizado por su coherencia y adecuación al contexto. En general, los resultados generados por el sistema son bastante precisos y tienden a optimizarse de manera iterativa con ajustes continuos del modelo.

Producción y agencia discursivas condicionadas por los sistemas de IA

A partir de estas condiciones sociotécnicas, y en articulación con el enfoque socio-cultural que orienta este trabajo, resulta necesario problematizar la participación del usuario en los entornos mediados por IA. En este contexto, la noción de agencia de los usuarios (Zavala, 2009, 2011) adquiere centralidad, en especial en ámbitos académicos, y puede entenderse como resultado de una participación situada, por medio de la cual interpretan, negocian e incluso cuestionan los posicionamientos que el sistema pone en circulación.

A partir de ello, puede interpretarse que los modelos de administración de infraestructura ofrecidos por empresas tecnológicas como Google o Microsoft a través de servicios como Google Cloud, Microsoft Azure, constituyen expresiones de poder corporativo (Crawford, 2021), en tanto consolidan discursos hegemónicos dictados por quienes controlan los inmensos volúmenes de datos. Estos modelos integran redes, servidores, sistemas operativos, entornos de ejecución y almacenamiento de datos, tanto en entornos locales como virtualizados parcial o completamente mediante aplicaciones gestionadas en la nube. Este tipo de infraestructura condiciona las capacidades de procesamiento y de modelado en función del corpus. Su funcionamiento por capas y niveles de gestión distribuye responsabilidades operativas sobre los componentes del sistema. Por ello, reconfigura el valor simbólico de los datos y los grados de autonomía del usuario en la producción discursiva. Asimismo, desde el enfoque sociotécnico las capas representan mediaciones semiótico-informáticas para la selección y transformación de los datos, jerarquizando al mismo tiempo los significados, por lo que no deben verse como herramientas computacionales neutras, objetivas o imparciales, ya que pueden reproducir y aumentar brechas estructurales existentes (Crawford, 2021).

Empresas como Open AI, Meta o Amazon se han centrado en la oferta de servicios de aplicaciones para usuarios finales y nuevas funcionalidades que mantienen el dominio de los flujos y el almacenamiento de datos. A medida que el control y la responsabilidad operativa del usuario sobre estos componentes son menores, hasta no requerir de su intervención, aumenta la delegación en el proveedor, privilegiando la simplicidad de uso frente a un mayor control corporativo (Crawford, 2021). Para visibilizar las relaciones de poder que atraviesan la producción discursiva mediada por IA, se requieren investigaciones empíricas que den cuenta de las implicancias

éticas y políticas, por ejemplo, en prácticas de uso de los usuarios o en el manejo corporativo de los datos y en las condiciones contractuales, gratuitas o de pago.

La siguiente y última sección presenta la propuesta teórica que concibe el lenguaje como sistema predictivo, orientada a los estudios del discurso mediado por IA desde un enfoque sociocultural de prácticas situadas.

EL LENGUAJE COMO SISTEMA PREDICTIVO PARA LOS ESTUDIOS DEL DISCURSO

La caracterización del lenguaje como sistema predictivo propuesta en este artículo supone una reformulación conceptual para comprender el lenguaje no solo como un objeto de análisis lingüístico o una práctica social, sino también como un sistema susceptible de modelación probabilística, capaz de anticipar elecciones discursivas en contextos específicos.

Perspectivas consolidadas de la lingüística y los estudios del discurso han permitido la comprensión estructural, semiótica, pragmática y social del lenguaje (Chomsky, 1969/1991; Halliday, 1982/2017; Street, 2004). No obstante, como se ha afirmado, las tecnologías predictivas de lenguaje basadas en IA introducen una dimensión socio-técnica adicional que reconfigura estos campos de estudio. En este nuevo escenario, el lenguaje no solo se interpreta o se usa, sino que también es anticipado, modelado y sugerido en tiempo real por sistemas que operan sobre grandes corpus de datos auténticos, de modo que la construcción discursiva ya no depende exclusivamente de los sujetos, sino también de algoritmos diseñados para asistir a los usuarios.

Por lo anterior, los marcos actuales resultan insuficientes para abordar las nuevas mediaciones socioinformáticas basadas en IA, pues los enfoques computacionales solo están centrados en aspectos técnicos y los tradicionales no incorporan las dimensiones sociotécnicas.

Esta propuesta plantea reconocer el lenguaje mediado por IA como traducción discursiva a términos numéricos derivada de la modelización algorítmica, además de incorporar la noción de práctica sociodiscursiva históricamente situada, en la que el significado depende del contexto y de los usos. De este modo, la producción discursiva puede entenderse como sistema de opciones cuya regularidad estadística es susceptible de modelación computacional.

Por otra parte, aunque los modelos de inferencia procesan datos auténticos que anticipan elecciones discursivas coherentes con el contexto sociocultural y la situación comunicativa específica del usuario, el significado no surge de manera exclusiva de las elecciones del hablante como plantea la perspectiva sistémico-funcional, sino de la interacción entre el humano como agente de la práctica social y la máquina como sistema de procesamiento, lo que requiere un marco integrado que permita comprender que los sistemas computacionales correlacionan, de forma textual, significado y contexto mediante ajustes estadísticos (clasificaciones, agrupaciones, tendencias), pero tales correlaciones remiten a prácticas sociales previas, por lo que la predicción computacional, y la respuesta automatizada, deben interpretarse como una inferencia semiótica social diseñada para asistir al usuario y priorizar sus intenciones comunicativas.

Cabe añadir que este escenario es opaco para la mayoría de los usuarios y da lugar a un fenómeno conocido como “caja negra” (Zemčík, 2026), ya que es imposible rastrear cómo los sistemas de autoaprendizaje llegan a sus resultados, incluso cuando su desempeño en la práctica parezca eficaz. Aunque los datos que se manejan en el sistema sean transparentes y controlados, la lógica interna del algoritmo permanece inaccesible debido a su enorme escala y complejidad. Esta problemática es relevante para los usuarios, ya que estos sistemas al operar mediante correlaciones probabilísticas no hacen explícitos sus fundamentos interpretativos. Además, por su alta eficacia es más fácil aceptar como válidas derivaciones estadísticas cuya procedencia o fundamento no son identificables con claridad. En este sentido, la producción de discursos sesgados o irrelevantes y la generación de conocimiento desvinculado de sus fuentes constituyen importantes desafíos académicos. Por ello, comprender los principios de la modelización algorítmica junto con el valor simbólico de los datos en la construcción discursiva asistida con IA resulta clave para entornos académicos.

La tabla presenta las perspectivas teóricas revisadas en diálogo con la propuesta semiótico-informática, incluyendo dimensiones, conceptos clave, categorías analíticas, características y ejemplos de ámbitos de aplicación que privilegia cada perspectiva, los cuales pueden utilizarse en futuras investigaciones empíricas para operar y validar esta propuesta, seleccionar corpus y diseñar estrategias metodológicas.

Su aporte consiste en proponer una articulación de la concepción sistémico-cognitiva del lenguaje predominante en el ámbito ingenieril con un enfoque sociocultural que, aunque resalta una agencia distribuida, busca destacar la centralidad de los usuarios y de sus prácticas como actores principales en los entornos digitales.

Tabla. Evolución de enfoques y propuesta de una nueva perspectiva semiótico-informática del discurso mediado por IA

Perspectiva teórica	Concepción del lenguaje	Dimensión analítica	Conceptos clave	Categorías analíticas	Características generales	Preguntas analíticas	Ámbitos de aplicación
Lingüística cartesiana (Chomsky, 1969/1991)	Capacidad cognitiva innata y sistema formal autónomo e independiente	Cognitivo-estructural	Competencia lingüística, facultad del lenguaje	Estructura profunda: léxico, sintaxis, semántica, estructura superficial: fonología y morfología	Foco en la naturaleza del lenguaje. Prioriza la estructura interna del lenguaje. Busca explicar los principios universales del uso y la adquisición del lenguaje	¿Qué estructura universal básica subyace a la producción lingüística? ¿Cómo está organizado el sistema lingüístico?	Procesamiento del lenguaje natural
Lingüística sistémico-funcional (Halliday, 1982/2017)	Sistema semiótico de opciones por el cual los hablantes construyen significados según sus intenciones comunicativas y contextos sociales	Semiótico-funcional	Metafunción ideacional, interpersonal y textual, cláusula, sistema, significado	Registro (campo, tenor y modo), transitividad, relaciones lógico-semánticas, modalidad, valoración, tema y rema, semántica discursiva, léxico-gramática y fonología	Teoría orientada a los usos del lenguaje. Analiza la relación entre forma, significado y contexto. El significado surge de las elecciones realizadas por los hablantes. Integra la estructura lingüística con las metafunciones del lenguaje	¿Qué marco actividad social configura el texto y dónde ocurre? ¿Qué significados construye el texto? ¿Quiénes participan, cuáles son sus relaciones y sus medios?	Análisis del discurso, pedagogía basada en géneros, traducción e interpretación, variación lingüística, enseñanza de lenguas

Nuevos estudios de literacidad (Street, 2004)	Lectura y escritura como actividades socioculturales e históricamente situadas que construyen realidades sociales e ideologías	Sociocultural	Sistemas de actividad, prácticas letradas, poder, identidades sociales, comunidades discursivas, ideologías y tecnologías	Contextos de uso, propósitos comunicativos, producción discursiva, relaciones de poder, roles de participación, normas y convenciones, tecnologías y medios	Predominan estudios de corte etnográfico y empíricos. Analiza la lectura y escritura como prácticas sociales situadas. Centrada en los sujetos, las comunidades y los contextos de uso. El significado se construye socialmente vinculado a estructuras de poder e ideologías	¿Cómo se lee y escribe en contextos específicos? ¿Quiénes participan y con qué propósitos? ¿Qué relaciones de poder, identidades e ideologías se construyen o reproducen?	Alfabetización académica, estudios de cultura digital y multimodalidad, literacidades emergentes y multiliteracidades, pedagogías críticas
Nueva semiótico-informática del discurso mediado por IA	Sistema parcialmente predecible mediante modelación algorítmica e inferencia semiótica, el cual está orientado a la asistencia del usuario. Actividad sociosemiótica e históricamente situada en la que los significados emergen de la interacción entre usuarios, modelos predictivos de IA y contextos sociotécnicos	Semiótico-informática y sociocultural	Sistema sociotécnico, práctica social situada, semiótica social, datos multimodales, lenguaje natural generado, mediaciones tecnológica, algorítmica y simbólica, poder corporativo y agencia distribuida	Contextos de uso, interacción humano-IA, <i>prompts</i> (datos de entrada multimodales), respuestas generadas, alineación contextual, patrones y regularidades discursivas, sesgos algorítmicos, infraestructura tecnológica, algoritmos y modelos predictivos, relaciones de poder y agencia	Propuesta teórica orientada a estudios del discurso mediado por IA. Analiza la interacción entre usuarios y sistemas sociotécnicos que generan lenguaje natural y priorizan las intenciones del usuario. La interacción reiterada tiende a estabilizar regularidades sociodiscursivas, pero también introduce sesgos. El significado y la interpretación del contexto surgen como inferencia semiótico-informática a partir de datos auténticos ajustados estadísticamente	¿Cómo interviene la predicción estadística en la producción de significados sociales y situados? ¿Qué papel desempeñan las intenciones del usuario en la interacción con el sistema sociotécnico de IA y la producción discursiva? ¿Qué patrones, regularidades y sesgos discursivos emergen de las interacciones?	Estudios del discurso mediado con IA, alfabetización crítica digital, alfabetización algorítmica, escritura asistida por IA, interacción educativa con <i>chatbots</i> , formación docente y uso institucional de IA, alfabetización informática y búsquedas con agentes de IA

Finalmente, sostenemos que los sistemas sociotécnicos de IA anticipan el lenguaje y, por tanto, moldean expectativas, orientan decisiones discursivas y condicionan horizontes interpretativos. Comprender este fenómeno exige una aproximación semiótico-informática, como la propuesta en este estudio, que articule las dimensiones técnicas y socioculturales sin subsumir una en la otra.

CONCLUSIONES

La propuesta de un marco teórico para explorar las condiciones contemporáneas de producción de conocimiento y las prácticas discursivas en entornos mediados por IA, así como sus implicaciones generales de uso, partió de la hipótesis de que la mediación algorítmica de los sistemas de IA introduce una lógica predictiva del lenguaje que reconfigura las formas de producción y circulación discursiva. Con base en ello, se plantean las bases de una perspectiva semiótico-informática orientada al estudio de los discursos mediados por IA.

El principal aporte conceptual al campo de los estudios del discurso consiste en proponer una vía de análisis para estudiar el lenguaje mediado por IA como un fenómeno discursivo, predictivo, sociotécnico y algorítmicamente mediado, desde una perspectiva semiótico-informática. Ello permite comprender que los modelos generativos operan mediante correlaciones estadísticas aprendidas a partir de grandes corpus y que, en determinados sistemas, sus procesos de ajuste pueden incorporar datos de uso, retroalimentación de usuarios y elementos contextuales, lo que pone de relieve el valor simbólico y económico que dichos datos pueden adquirir para las empresas tecnológicas. Asimismo, esta perspectiva permite explicar que las respuestas generadas por IA son resultado de procesos automatizados de inferencia probabilística.

Respecto de las implicaciones generales de la interacción con asistentes de IA que sugieren, completan y reformulan el lenguaje, resulta relevante la noción de agencia distribuida (Kudina y Van de Poel, 2024), ya que las elecciones discursivas de los usuarios se producen en interacción con arquitecturas algorítmicas y dependen también de sus motivaciones, intenciones y criterios éticos de uso, entre otros factores.

En el ámbito académico, este escenario desplaza el problema tradicional del plagio textual (Castro y Sánchez, 2021) hacia una cuestión más compleja, relacionada con la trazabilidad y la procedencia epistemológica de los contenidos generados mediante asistentes de IA. Dado que estos sistemas no siempre permiten identificar con precisión las fuentes de la información que producen y que las referencias proporcionadas pueden ser incompletas o incorrectas, se dificulta la atribución de la autoría y de la autoridad intelectual de los contenidos. En consecuencia, las respuestas generadas por IA no deberían considerarse por sí mismas fuentes académicas equivalentes a las fuentes originales, por lo que resulta necesario verificar y citar directamente estas últimas cuando sean identificables. Asimismo, es indispensable reconocer otros límites de la escritura mediada por IA, en particular los sesgos vinculados con los datos y con los procesos de entrenamiento, clasificación, ajuste y personalización. Frente a estas problemáticas, resulta necesario establecer estándares de uso responsable y transparente de la IA en las universidades, dirigidos tanto a estudiantes y docentes como a las instituciones y empresas tecnológicas involucradas.

Es preciso señalar que este trabajo constituye una aproximación teórica inicial que requiere contrastación empírica, ya que, debido a su enfoque, no se analizaron corpus de interacciones humano-IA ni los efectos discursivos, cognitivos y educativos de la mediación algorítmica en aulas o instituciones. Estas limitaciones abren futuras líneas de investigación orientadas a evaluar el modelo propuesto y profundizar en las implicaciones educativas y éticas del uso de la IA mediante estudios empíricos en el aula, análisis de la escritura asistida por IA e intervenciones pedagógicas centradas en la alfabetización digital y algorítmica.

Finalmente, este trabajo contribuye a la actualización de los marcos y las pedagogías vinculados con las literacidades digitales, al proponer una comprensión crítica del lenguaje mediado por tecnologías de IA. Asimismo, plantea la necesidad de contar con políticas institucionales para el uso de estas tecnologías, con el propósito de fortalecer la integridad epistemológica de las prácticas académicas contemporáneas. Para ello, se requiere diseñar protocolos universitarios que orienten y acompañen las prácticas docentes y estudiantiles frente al uso de la IA generativa y que sean resultado de la participación de los distintos actores involucrados.

REFERENCIAS BIBLIOGRÁFICAS

- Atkinson-Abutridy, J. (2023a). *Analítica textual: introducción a la ciencia y aplicación del análisis de información no estructurada*. Marcombo.
- Atkinson-Abutridy, J. (2023b). *Grandes modelos de lenguaje: conceptos, técnicas y aplicaciones*. Marcombo.
- Bearman, M., Ryan, J. y Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, vol. 86, núm. 2, pp. 369-385. <https://doi.org/10.1007/s10734-022-00937-2>
- Beltrán, N. y Rodríguez, E. (2020). Procesamiento del lenguaje natural (PLN)-GPT-3: aplicación en la ingeniería de Software. *Tecnología, Investigación y Academia*, vol. 8, núm. 1, pp. 18-37. <https://revistas.udistrital.edu.co/index.php/tia/article/view/17323>
- Castro, M. y Sánchez, M. (2021). Uso de los recursos de cita y referencia en la escritura de la tesis de posgrado y su relación con el plagio textual. *Diálogos sobre Educación. Temas Actuales en Investigación Educativa*, vol. 12, núm. 23, pp. 1-20.
- Chomsky, N. (1969/1991). *Lingüística cartesiana: un capítulo de la historia del pensamiento racionalista*. Editorial Gredos.
- Cortez, A., Vega, H. y Pariona, J. (2009). Procesamiento de lenguaje natural. *Revista de Investigación de Sistemas e Informática*, vol. 6, núm. 2, pp. 45-54. <https://link.gale.com/apps/doc/A309314630/IFME?u=anon~d4f0080c&sid=googleScholar&xid=0a11cd9c>
- Crawford, K. (2021). Conclusion. Power. En *The Atlas of AI* (pp. 211-227). Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t.10>
- Crompton, H. y Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, vol. 20, núm. 1, pp. 1-22. <https://doi.org/10.1186/s41239-023-00392-8>
- Delso, A., Carvajal, M. y Corral, D. (2025). La evolución del procesamiento del lenguaje natural y su influencia en la inteligencia artificial: una revisión y líneas de investigación futura. *European Public y Social Innovation Review*, vol. 10, pp.1-23. <https://doi.org/10.31637/epsir-2025-782>
- Dietz, G., King, J., Beason, J., Tarrow, M., Hilliard, A. y Shapiro, R. (2022). *ARtonomous: Introducing middle school students to reinforcement learning through virtual robotics*. Proceedings of the 21st Annual ACM Interaction Design and Children Conference, pp. 430-441. <https://doi.org/10.1145/3501712.35297>
- Domínguez, A. (2002). Lingüística computacional: un esbozo. *Boletín de Lingüística*, vol. 18, pp. 114-119. <https://biblat.unam.mx/hevila/Boletindelinguistica/2002/no18/5.pdf>
- Estrada, O., Rincón, E. y Mena, J. (2026). Inteligencia artificial y comunicación: mediación crítica y educ comunicativa. *Index. Comunicación: Revista Científica en el Ámbito de la Comunicación Aplicada*, vol. 16, núm. 1, pp. 13-36. <https://doi.org/10.62008/ixc/16/01Artifi>
- Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for Social Research*. Routledge.
- Forero-Corba, W. y Bennasar, F. (2024). Técnicas y aplicaciones del machine learning e inteligencia artificial en educación: una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, vol. 27, núm. 1, pp. 209-253.

<https://doi.org/10.5944/ried.27.1.37491>

- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos. Revista de Ciencias Sociales*, vol. 11, núm. 2, pp.1-17. <https://doi.org/10.17502/mrcs.v11i2.710>.
- Fúquene, H. (2024). Procesamiento de lenguaje natural, los transformers y los bots conversacionales. *XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan*, vol. 12, pp. 151-160. <https://doi.org/10.29057/xikua.v12iEspecial.12904>
- García-Peñalvo, F., Llorens-Largo, F. y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, vol. 27, núm. 1, pp. 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- García-Peñalvo, F. y Vásquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. *IJIMAI*, vol. 8, núm. 4, pp. 7-16. <https://doi.org/10.9781/ijimai.2023.07.006>
- Giraldo, A. y Orozco, A. (2023). Evolución del procesamiento natural del lenguaje. *TecnoLógicas*, vol. 26, núm. 56, pp.1-3. <https://doi.org/10.22430/22565337.2687>
- Goodfellow, I., Bengio, Y. y Courville, A. (2017). *Deep learning* (vol. 1). MIT Press Cambridge.
- Halliday, M. (1982/2017). *Lenguaje como semiótica social: la interpretación social del lenguaje y del significado*. Fondo de Cultura Económica.
- Henao-Villa, C., García-Arango, D., Aguirre-Mesa, E., González-García, A., Bracho-Aconcha, R., Solorzano-Movilla, J. y Arboleda-Lopez, A. (2017). Multidisciplinariedad, interdisciplinariedad y transdisciplinariedad en la formación para la investigación en ingeniería. *Revista Lasallista de Investigación*, vol. 14, núm. 1, pp. 179-197. <https://doi.org/10.22507/rli.v14n1a16>
- Jones, R. y Hafner, C. (2021). Online cultures and intercultural communication. En R. Jones y C. Hafner (eds.). *Understanding digital literacies* (pp. 161-182). Routledge. <https://doi.org/10.4324/9781003177647-10>
- Jones, R. y Hafner, C. (2012). *Understanding digital literacies*. Routledge.
- Kaur, D., Uslu, S., Rittichier, K. y Durrezi, A. (2022). Trustworthy artificial intelligence: a review. *ACM computing surveys (CSUR)*, vol. 55, núm. 2, pp. 1-38. <https://doi.org/10.1145/349120>
- Kosmyna, N., Hauptmann, E., Yuan, Y., Situ, J., Liao, X., Beresnitzky, A. y Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, vol. 4, pp. 1-206.
- Kudina, O. y Van de Poel, I. (2024). A sociotechnical system perspective on AI. *Minds and Machines*, vol. 34, núm. 21, pp. 1-9. <https://doi.org/10.1007/s11023-024-09680-2>
- McCarthy, J., Minsky, M., Rochester, N. y Shannon, C. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine*, vol. 27, núm. 4, pp. 12-12. <https://doi.org/10.1609/ai-mag.v27i4.1904>
- Méndez-Ochaita, M. F. (2022). Multiliteracidades en la era digital: conceptos clave desde los estudios de literacidad y multimodalidad. *Revista Electrónica*

- Leer, Escribir y Descubrir, vol. 1, núm. 11, pp. 11-22. <https://digitalcommons.fiu.edu/record/19647?v=pdf>
- Méndez-Ochaita, M. y Hernández-Ramírez, L. (2026). Variantes en perfiles de literacidad digital: hacia una práctica crítica de estudiantes universitarios. *Folios*, núm. 63, pp. 99-116. <https://doi.org/10.17227/folios.63-22014>
- Ocaña-Fernández, Y., Valenzuela-Fernández, L. y Garro-Aburto, L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, vol. 7, núm. 2, pp. 536-568. <https://doi.org/10.20511/pyr2019.v7n2.274>
- Poole, D., Mackworth, A. y Goebel, R. (1998). *Computational intelligence: A logical approach*. Oxford University Press.
- Porras, J., Florencia-Juarez, R., Rivera, G. y García, V. (2018). Interfaz de lenguaje natural para consultar cubos multidimensionales utilizando procesamiento analítico en línea. *Research in Computing Science*, vol. 147, núm. 6, pp. 153-165. <http://cathi.uacj.mx/20.500.11961/7984>
- Salas-Pilco, S. y Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, vol. 19, núm. 1, pp. 1-20. <https://doi.org/10.1186/s41239-022-00326-w>
- Sanabria-Navarro, J., Silveira-Pérez, Y., Pérez-Bravo, D. y Cortina-Núñez, M. (2023). Incidencias de la inteligencia artificial en la educación contemporánea. *Comunicar: Revista Científica de Comunicación y Educación*, núm. 77, pp. 97-107. <https://doi.org/10.3916/C77-2023-08>
- Sánchez, M. (2021). Las tecnologías de la información y su relación con la escritura. *Revista Eduweb*, vol. 15, núm. 3, pp. 62-68. <https://doi.org/10.46502/issn.1856-7576/2021.15.03.5>
- Schaub, L. (2020). La industria del lenguaje en la era del dato. *Ábaco*, núm. 103, pp. 82-89. <https://www.jstor.org/stable/27135841>
- Segura, M., Mello, J. y Hernández, A. (2022). Machine learning prediction of university student dropout: Does preference play a key role? *Mathematics*, vol. 10, núm. 18, pp. 1-20. <https://doi.org/10.3390/math10183359>
- Soto, V., Quintero, V., Callejas, O. y Rodríguez, G. (2024). Transformando la ingeniería con la inteligencia artificial y el aprendizaje automático, retos y oportunidades. *Innovación y Desarrollo Tecnológico*, vol. 16, núm. 4, pp. 1503-1522. https://iydt.wordpress.com/wp-content/uploads/2024/08/4_14_transformando-la-ingenieria-con-la-inteligencia-artificial-y-el-aprendizaje-automatico-retos-y-oportunidades.pdf
- Street, B. (2009). Perspectivas etnográficas y políticas sobre cultura escrita: el poder de nombrar y definir. En J. Kalman y B. Street (eds.). *Lectura, escritura y matemáticas como prácticas sociales. Diálogos con América Latina* (pp. 84-111). Siglo XXI Editores.
- Street, B. (2004). Los nuevos estudios de literacidad. En V. Zavala, M. Niño-Murcia y P. Ames (eds.). *Escritura y sociedad. Nuevas perspectivas teóricas y etnográficas* (pp. 81-107). Red para el Desarrollo de las Ciencias Sociales en el Perú.
- Taha, S., Shihab, R. y Sadik, M. (2018). Studying of educational data mining techniques. *International Journal of Advanced Research in Science, Engineering and Technology*, vol. 5, núm. 5, pp. 5742-5750.
- Van Dijk, T. (2019). *El discurso como interacción social*. Gedisa.

- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R. y Funtowicz, M. (2020). *Transformers: State-of-the-art natural language processing*. Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, pp. 38-45.
- Young, T., Hazarika, D., Poria, S. y Cambria, E. (2018). Recent trends in deep learning based natural language processing [review article]. *IEEE Computational Intelligence Magazine*, vol. 13, núm. 3, pp. 55-75. <https://doi.org/10.1109/MCI.2018.2840738>
- Zavala, V. (2011). La escritura académica y la agencia de los sujetos. *Cuadernos Comillas*, vol. 1, núm. 56, pp. 52-66.
- Zavala, V. (2009). ¿Quién está diciendo eso?: literacidad académica, identidad y poder en la educación superior. En J. Kalman y B. Street (eds.). *Lectura, escritura y matemáticas como prácticas sociales. Diálogos con América Latina* (pp. 348-363). Siglo XXI Editores.
- Zawacki-Richter, O., Marín, V., Bond, M. y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, vol. 16, núm. 39, pp. 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zemčik, T. (2026). A brief history of definitions of artificial intelligence definitions: From the great bombe to black boxes. *Pedagogika*, vol. 75, núm. 4, pp. 309-340. <https://doi.org/10.14712/23362189.2025.5009>