

Procesamiento del Lenguaje Natural

Introducción al PLN

Clase 1

Maestría en Educación en Inteligencia Artificial y Entornos Virtuales

La excelencia no se improvisa



Introducción

Esta clase introduce el Procesamiento de Lenguaje Natural (PLN), un campo interdisciplinario que combina la lingüística, la inteligencia artificial y la informática para desarrollar sistemas capaces de comprender, interpretar y generar lenguaje humano. Se abordarán conceptos fundamentales como la tokenización, el análisis sintáctico, la semántica y la pragmática, esenciales para la construcción de modelos lingüísticos avanzados. Además, se hace un recorrido por la evolución del PLN, desde los primeros intentos basados en reglas hasta la revolución impulsada por redes neuronales y modelos de lenguaje como GPT y BERT. También se destacan diversas aplicaciones del PLN en sectores como la medicina, la educación, las finanzas y el entretenimiento.

En términos de aprendizaje, los estudiantes desarrollarán una comprensión sólida de los principios básicos del PLN y su importancia en el ámbito tecnológico actual. Analizarán los conceptos y técnicas de PLN en ambientes educativos. Adquirirán conocimientos sobre la morfología, la sintaxis y la semántica desde una perspectiva computacional, explorando cómo estos conceptos se aplican en tareas como la traducción automática, la generación de texto y el análisis de datos lingüísticos.

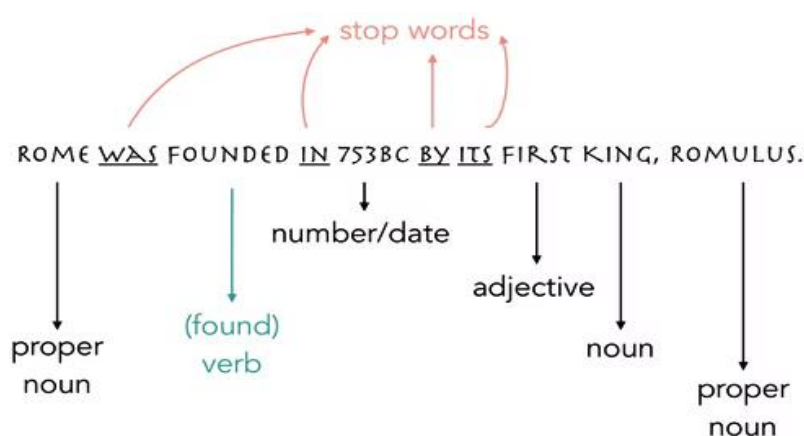
Clase 1

1.Introducción al PLN

1.1. Conceptos relacionados al PLN

El Procesamiento de Lenguaje Natural (PLN) es una disciplina interdisciplinaria que combina la lingüística, la inteligencia artificial y las ciencias de la computación para permitir que las máquinas comprendan, interpreten y generen lenguaje humano. Entre los conceptos clave del PLN se encuentran la tokenización (Figura 1), el análisis sintáctico, la semántica y la pragmática. La tokenización es el proceso de dividir un texto en unidades más pequeñas, como palabras o frases, para su análisis (Jurafsky y Martin, 2021). Por ejemplo, la frase "El gato corre" se tokeniza en ["El", "gato", "corre"].

Figura 1. Tokenización



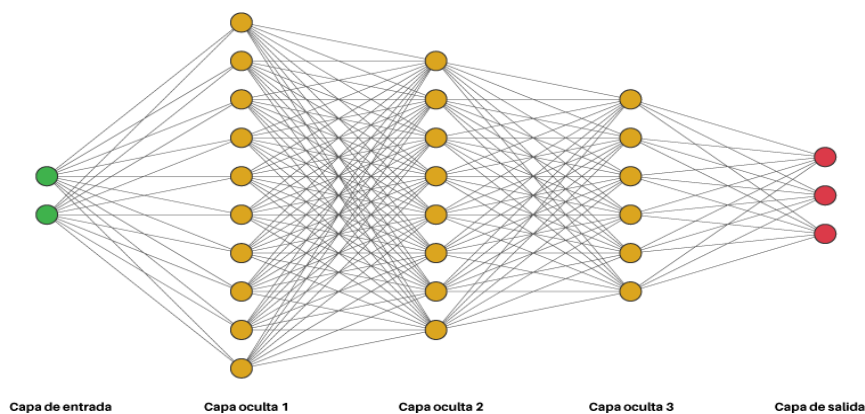
Nota: Tomado de Peláez (2022)

El análisis sintáctico se refiere a la estructura gramatical de una oración, identificando sujetos, verbos y objetos. Herramientas como los analizadores de dependencias permiten visualizar estas relaciones (Manning y Schütze, 1999). Por otro lado, la semántica se enfoca en el significado de las palabras y

oraciones. Un ejemplo es la resolución de ambigüedades, como distinguir si "banco" se refiere a una institución financiera o a un asiento (Cambria y White, 2014). Finalmente, la pragmática estudia el contexto en el que se usa el lenguaje, como la ironía o el sarcasmo, que son desafíos complejos para los sistemas de PLN (Bender y Koller, 2020).

Otro concepto importante es el de los modelos de lenguaje, como los basados en redes neuronales, por ejemplo, GPT o BERT, que han revolucionado el campo al permitir que las máquinas generen texto coherente y contextualmente relevante (Vaswani et al., 2017). Estos modelos se entrenan con grandes volúmenes de datos textuales para predecir la probabilidad de secuencias de palabras. El PLN es un campo interdisciplinario que busca cerrar la brecha entre el lenguaje humano y la comprensión computacional.

Figura 2. Red Neuronal



Nota: Tomado de Codificando Bits (2025)

1.2. Historia de la evolución del PLN

El PLN tiene sus raíces en la década de 1950, cuando Alan Turing propuso la idea de máquinas capaces de entender y generar lenguaje (Figura 3) (Turing, 1950). En 1954, el experimento Georgetown-IBM marcó un hito al traducir automáticamente más de 60 oraciones del ruso al inglés (Hutchins, 2005). Sin embargo, los sistemas de esa época se basaban en reglas lingüísticas predefinidas, lo que limitaba su capacidad para manejar la complejidad del lenguaje natural.

Figura 3. Classic Turing Machine

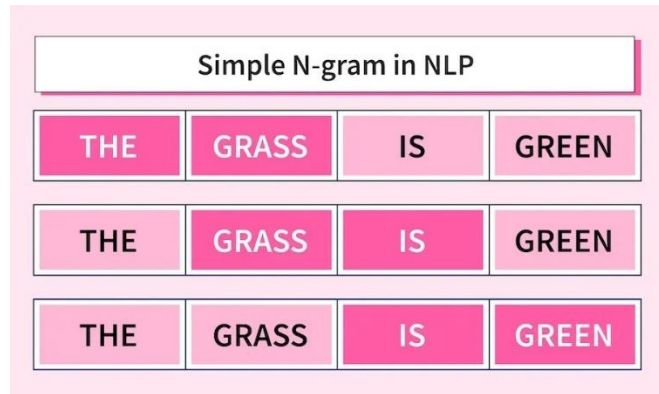


Nota: Tomado de Fishwick (2013)

En las décadas de 1980 y 1990, el enfoque cambió hacia métodos estadísticos, donde los sistemas aprendían patrones a partir de grandes corpus de texto. Por ejemplo, los modelos de n-gramas (Figura

4) permitían predecir la siguiente palabra en una secuencia basándose en la frecuencia de ocurrencia (Manning y Schütze, 1999). Este enfoque impulsó aplicaciones como los correctores ortográficos y los motores de búsqueda.

Figura 4. What is N-Gram and How does it work?



Nota: Tomado de Relinns Technologies (2025)

La llegada del aprendizaje profundo en la década de 2010 revolucionó el PLN. Modelos como Word2Vec (Mikolov et al., 2013) introdujeron la idea de representaciones vectoriales de palabras, capturando relaciones semánticas como "rey - hombre + mujer = reina". Posteriormente, arquitecturas como Transformer (Vaswani et al., 2017) y modelos preentrenados como GPT y BERT permitieron avances sin precedentes en tareas como traducción automática, resumen de texto y generación de respuestas. Hoy, el PLN es una piedra angular de la inteligencia artificial, con aplicaciones que impactan múltiples industrias.

Para conocer más detalles sobre la apasionante evolución del PLN, recomendamos revisar esta versión ampliada: <https://www.dlsi.ua.es/~japerez/materials/historia/>.

1.3. Aplicaciones con PLN en diversos ámbitos

El PLN tiene aplicaciones en una amplia gama de campos, desde la medicina hasta el entretenimiento. En el ámbito médico, se utiliza para extraer información de historiales clínicos, lo que facilita el diagnóstico y la investigación. Por ejemplo, sistemas como IBM Watson Health (Figura 5) analizan textos médicos para identificar patrones y sugerir tratamientos (Topol, 2019).

Figura 5. IBM Watson Health



Nota: Tomado de MedImaging (2017)

En el ámbito educativo, el PLN se utiliza para personalizar el aprendizaje y mejorar la interacción entre estudiantes y plataformas de aprendizaje. Sistemas como Grammarly, que emplean PLN para proporcionar sugerencias de gramática y estilo, son ampliamente utilizados por estudiantes y profesionales para mejorar la calidad de sus escritos.

Además, plataformas de aprendizaje en línea, como Duolingo, utilizan PLN para adaptar el contenido educativo a las necesidades individuales de los usuarios. Por ejemplo, el análisis de respuestas y errores de los estudiantes permite ajustar dinámicamente las lecciones para optimizar el aprendizaje.

Otro ejemplo innovador es el uso de sistemas de análisis de textos para evaluar automáticamente ensayos y tareas escritas, ofreciendo retroalimentación inmediata y reduciendo la carga de trabajo de los educadores.

En el sector financiero, el PLN se emplea para analizar sentimientos en redes sociales y predecir tendencias del mercado. Herramientas como Sentiment Analysis permiten a los inversores tomar decisiones basadas en la opinión pública (Liu, 2015). Además, los chatbots impulsados por PLN, como los de bancos, ofrecen asistencia automatizada a los clientes.

En el ámbito legal, el PLN ayuda a revisar contratos y documentos legales, identificando cláusulas clave y reduciendo el tiempo de análisis. Por ejemplo, plataformas como Kira Systems automatizan la extracción de información relevante (Susskind, 2017). En el entretenimiento, asistentes virtuales como Siri o Alexa utilizan PLN para entender y responder a comandos de voz, mientras que plataformas como Netflix lo emplean para generar recomendaciones personalizadas (Gomez-Uribe & Hunt, 2016).

Otra aplicación destacada es la traducción automática, donde herramientas como Google Translate han mejorado significativamente gracias a modelos neuronales (Wu et al., 2016). Estas aplicaciones demuestran cómo el PLN está transformando industrias y mejorando la interacción humano-máquina.

1.4. Panorama actual del PLN

El panorama actual del PLN está dominado por modelos de lenguaje preentrenados y arquitecturas basadas en Transformer. Modelos como GPT-4, BERT y T5 han establecido nuevos estándares en tareas como generación de texto, traducción y comprensión de contexto (Brown et al., 2020; Devlin et al., 2019). Estos modelos se entrenan con enormes cantidades de datos y requieren un poder computacional significativo, lo que ha llevado a colaboraciones entre empresas tecnológicas y académicas.

Uno de los desafíos actuales es la ética en el PLN. Los modelos pueden perpetuar sesgos presentes en los datos de entrenamiento, lo que ha llevado a un mayor enfoque en la transparencia y la equidad (Bender et al., 2021). Por ejemplo, se están desarrollando técnicas para detectar y mitigar sesgos en aplicaciones como reclutamiento automatizado o vigilancia en redes sociales.

Además, el PLN está impulsando la personalización en servicios digitales. Plataformas como Spotify y Amazon utilizan modelos de lenguaje para entender las preferencias de los usuarios y ofrecer recomendaciones precisas (Gomez-Uribe y Hunt, 2016). En el ámbito de la investigación, el PLN está facilitando el análisis de grandes volúmenes de literatura científica, acelerando descubrimientos en campos como la medicina y la climatología (Topol, 2019).

En resumen, el PLN vive una era de rápido avance, con aplicaciones cada vez más sofisticadas y un impacto creciente en la sociedad. Sin embargo, también enfrenta desafíos éticos y técnicos que requieren atención continua. En esta sección se ha proporcionado un enfoque al PLN desde un punto de vista más bien académico. En el siguiente link se encuentra un video con un punto de vista de corte

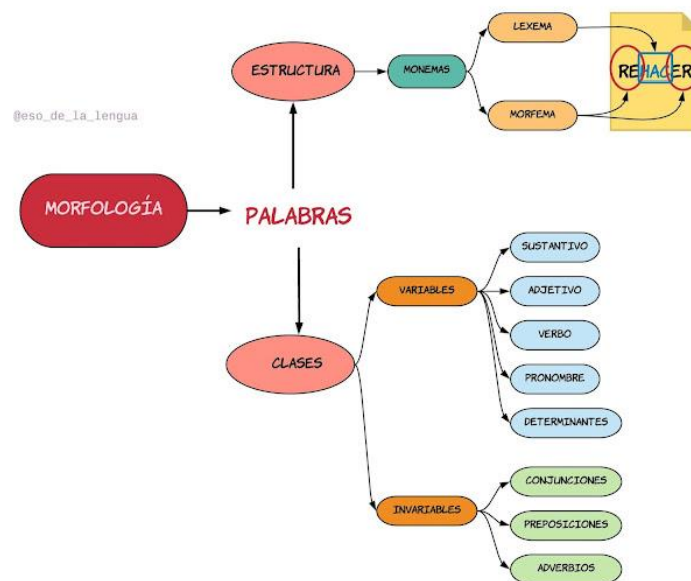
un poco más aplicado, que vale la pena ver para establecer comparaciones: <https://youtu.be/fLvJ8VdHLA0?si=gJ52d7-0OodsC7zt>.

2. Introducción a los conceptos fundamentales de lingüística computacional

2.1. Morfología en PLN

La morfología es una rama de la lingüística que estudia la estructura interna de las palabras y cómo se forman a partir de unidades más pequeñas llamadas morfemas (Figura 6). En el Procesamiento de Lenguaje Natural (PLN), la morfología juega un papel crucial en tareas como: el análisis de palabras, la generación de texto y la traducción automática. Este ensayo explora los conceptos clave de la morfología en PLN, sus aplicaciones y los desafíos que presenta.

Figura 6. Morfología



Nota: Tomado de Rodríguez et al. (2018)

Conceptos clave de la morfología

Un morfema (Figura 7) es la unidad mínima de significado en una lengua. Los morfemas pueden ser libres (palabras independientes como "gato") o ligados (prefijos, sufijos o infijos que no pueden existir por sí solos, como "-s" en "gatos"). Esta se divide en dos áreas principales: morfología flexiva y morfología derivacional. Por un lado, la morfología flexiva se ocupa de las variaciones de las palabras para expresar categorías gramaticales como el número, el género o el tiempo verbal (por ejemplo, "corro" vs. "corrió"). Por otro lado, la morfología derivacional se enfoca en la creación de nuevas palabras a partir de morfemas (por ejemplo, "feliz" → "felicidad") (Jurafsky y Martin, 2021).

Figura 7. Clases de Morfemas

CLASES DE MORFEMAS				
sub	mar	in	o	s
Morfema prefijo o derivativo	Morfema raíz o lexema	Morfema sufijo o derivativo	Morfemas flexivos. Indican accidentes gramaticales en el:	
Son generadores de palabras: • <i>subterráneo</i> • <i>suburbio</i> • <i>suboficial</i>	Es la parte principal del significado base de la palabra	Son generadores de palabras: • <i>azulino</i> • <i>taurino</i> • <i>caprino</i>	• género	} EN SUSTANTIVO
			• número	
			• persona	} EN VERBO
			• tiempo	
			• aspecto	
			• modo	

Nota: Tomado de Chipana (2014)

En PLN, el análisis morfológico implica descomponer las palabras en sus morfemas constituyentes. Por ejemplo, la palabra "inesperadamente" se puede descomponer en los morfemas "in-", "esper-", "-ada", y "-mente". Este proceso es esencial para tareas como la lematización, que consiste en reducir una palabra a su forma base o lema (por ejemplo, "corriendo" → "correr").

Aplicaciones de la morfología en PLN

Una de las aplicaciones más importantes de la morfología en PLN es la lematización. Los lematizadores son herramientas que utilizan reglas morfológicas para identificar la forma canónica de una palabra. Por ejemplo, el lematizador de SpaCy puede convertir "corriendo" en "correr" y "gatos" en "gato" (Honnibal y Montani, 2017). Esta técnica es fundamental en motores de búsqueda y sistemas de recuperación de información, donde es necesario normalizar las palabras para mejorar la precisión de los resultados.

Otra aplicación es la generación morfológica, que consiste en crear formas flexionadas o derivadas de una palabra. Por ejemplo, un sistema de traducción automática debe generar correctamente las formas plurales o los tiempos verbales en el idioma de destino. Los modelos basados en reglas y los modelos estadísticos han sido ampliamente utilizados para esta tarea, aunque los enfoques basados en redes neuronales, como los Transformers, han demostrado ser más efectivos (Wu et al., 2016).

Desafíos de la morfología en PLN

Uno de los principales desafíos en el análisis morfológico es la ambigüedad morfológica. Algunas palabras pueden tener múltiples interpretaciones dependiendo del contexto. Por ejemplo, la palabra "banco" puede referirse a una institución financiera o a un asiento, y su morfología no siempre resuelve esta ambigüedad. Los sistemas de PLN deben combinar el análisis morfológico con el contexto semántico para tomar decisiones precisas (Manning y Schütze, 1999).

Otro desafío es la diversidad lingüística. Las lenguas tienen estructuras morfológicas muy diferentes. Por ejemplo, el español es una lengua flexiva, donde las palabras suelen tener muchas formas flexionadas, mientras que el chino es una lengua aislante, donde las palabras tienden a ser invariables. Los sistemas de PLN deben adaptarse a estas diferencias para funcionar eficazmente en múltiples idiomas (Bender, 2011).

La morfología es un componente esencial del PLN que permite a las máquinas entender y generar palabras de manera precisa. Aunque existen desafíos como la ambigüedad morfológica y la diversidad

lingüística, los avances en modelos neuronales y técnicas de aprendizaje automático están mejorando significativamente las capacidades de los sistemas de PLN en este ámbito.

2.2. Sintaxis en PLN

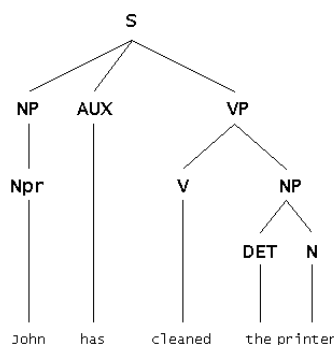
La sintaxis es el estudio de las reglas que gobiernan la estructura de las oraciones en un lenguaje. En el Procesamiento de Lenguaje Natural (PLN), el análisis sintáctico es fundamental para tareas como la traducción automática, la generación de texto y la comprensión del lenguaje. Este ensayo explora los conceptos clave de la sintaxis en PLN, sus aplicaciones y los desafíos que enfrenta.

Conceptos clave de la sintaxis

La sintaxis se ocupa de cómo las palabras se combinan para formar oraciones gramaticalmente correctas. Un concepto central es el de constituyentes, que son grupos de palabras que funcionan como una unidad dentro de una oración. Por ejemplo, en la oración "El gato negro corre rápido", "El gato negro" es un constituyente que funciona como sujeto (Carnie, 2021).

El análisis sintáctico en PLN se realiza mediante árboles de análisis, que representan la estructura jerárquica de una oración (Figura 8). Existen dos tipos principales de árboles: árboles de constituyentes y árboles de dependencias. Los árboles de constituyentes dividen la oración en frases y subfrases, mientras que los árboles de dependencias muestran las relaciones entre palabras individuales (Jurafsky y Martin, 2021). Por ejemplo, en la oración "El gato come pescado", el árbol de dependencias indicaría que "gato" es el sujeto de "come" y "pescado" es el objeto.

Figura 8. Árbol Sintáctico



Nota: Tomado de Moreno (2000)

Aplicaciones de la sintaxis en PLN

Una de las aplicaciones más importantes de la sintaxis en PLN es la traducción automática. Los sistemas de traducción deben analizar la estructura sintáctica de una oración en el idioma de origen y generar una oración con una estructura equivalente en el idioma de destino. Por ejemplo, el sistema de Google Translate utiliza modelos basados en Transformers para realizar análisis sintáctico y generar traducciones precisas (Wu et al., 2016).

Otra aplicación es la generación de texto, donde los sistemas deben producir oraciones gramaticalmente correctas y coherentes. Los modelos de lenguaje como GPT-4 utilizan información sintáctica para predecir la siguiente palabra en una secuencia, lo que les permite generar texto fluido y contextualmente relevante (Brown et al., 2020).

Desafíos de la sintaxis en PLN

Uno de los principales desafíos en el análisis sintáctico es la ambigüedad estructural. Algunas oraciones pueden tener múltiples interpretaciones sintácticas dependiendo del contexto. Por ejemplo, la oración "Vi al hombre con el telescopio" (Figura 9) puede interpretarse de dos maneras: el hombre tiene un telescopio o el hablante usó un telescopio para ver al hombre. Los sistemas de PLN deben utilizar información contextual para resolver estas ambigüedades (Manning y Schütze, 1999).

Figura 9. Vi al hombre con el telescopio



Nota: Imagen generada con DALL-E (2025)

Otro desafío es la diversidad lingüística. Las lenguas tienen estructuras sintácticas muy diferentes. Por ejemplo, el inglés sigue un orden básico de sujeto-verbo-objeto (SVO), mientras que el japonés sigue un orden de sujeto-objeto-verbo (SOV). Los sistemas de PLN deben adaptarse a estas diferencias para funcionar eficazmente en múltiples idiomas (Bender, 2011).

En este sentido, la sintaxis es un componente esencial del PLN que permite a las máquinas entender y generar oraciones gramaticalmente correctas. Aunque existen desafíos como la ambigüedad estructural (Figura 10) y la diversidad lingüística, los avances en modelos neuronales y técnicas de aprendizaje automático están mejorando significativamente las capacidades de los sistemas de PLN en este ámbito.

Figura 10. Ambigüedad



Nota: Tomado de Bescós (2020)

2.3. Semántica en PLN

La semántica es el estudio del significado de las palabras, frases y oraciones en un lenguaje. En el Procesamiento de Lenguaje Natural (PLN), el análisis semántico es crucial para tareas como la comprensión del lenguaje, la respuesta a preguntas y la generación de texto. Este ensayo explora los conceptos clave de la semántica en PLN, sus aplicaciones y los desafíos que enfrenta.

Conceptos clave de la semántica

La semántica se ocupa de cómo las palabras y las oraciones adquieren significado. Un concepto central es el de representación semántica, que es una formalización del significado en un formato que las máquinas pueden procesar. Por ejemplo, la oración "El gato está sobre la mesa" puede representarse como una relación entre "gato" y "mesa" con el predicado "sobre" (Jurafsky y Martin, 2021).

El análisis semántico en PLN se realiza mediante técnicas como el análisis de roles semánticos y la desambiguación de sentido de palabras. El análisis de roles semánticos identifica los papeles que juegan las palabras en una oración, como agente, paciente o instrumento. Por ejemplo, en la oración "Juan abrió la puerta con una llave", "Juan" es el agente, "puerta" es el paciente y "llave" es el instrumento (Palmer et al., 2010). La desambiguación de sentido de palabras, por otro lado, consiste en determinar el significado correcto de una palabra en un contexto dado. Por ejemplo, la palabra "banco" puede referirse a una institución financiera o a un asiento, y los sistemas de PLN deben utilizar el contexto para resolver esta ambigüedad (Navigli, 2009).

Aplicaciones de la semántica en PLN

Una de las aplicaciones más importantes de la semántica en PLN es la respuesta a preguntas. Los sistemas de respuesta a preguntas deben entender el significado de una pregunta y encontrar la información relevante en un texto. Por ejemplo, el sistema IBM Watson utiliza técnicas de análisis semántico para responder preguntas complejas en dominios específicos, como la medicina (Ferrucci et al., 2010).

Otra aplicación es la generación de texto, donde los sistemas deben producir texto que no solo sea gramaticalmente correcto, sino también semánticamente coherente. Los modelos de lenguaje como GPT-4 utilizan información semántica para predecir la siguiente palabra en una secuencia, lo que les permite generar texto fluido y contextualmente relevante (Brown et al., 2020).

Desafíos de la semántica en PLN

Uno de los principales desafíos en el análisis semántico es la ambigüedad semántica. Algunas palabras y oraciones pueden tener múltiples significados dependiendo del contexto. Por ejemplo, la oración "Vi al hombre con el telescopio" puede interpretarse de dos maneras: el hombre tiene un telescopio o el hablante usó un telescopio para ver al hombre. Los sistemas de PLN deben utilizar información contextual para resolver estas ambigüedades (Manning y Schütze, 1999).

Otro desafío es la comprensión del contexto. El significado de una palabra o frase puede depender de información que no está explícita en el texto. Por ejemplo, la frase "Está lloviendo" puede tener diferentes implicaciones dependiendo de si el hablante está planeando un picnic o regando un jardín. Los sistemas de PLN deben ser capaces de inferir este tipo de información contextual para entender completamente el significado (Bender y Koller, 2020).

La semántica es un componente esencial del PLN que permite a las máquinas entender y generar texto con significado preciso. Aunque existen desafíos como la ambigüedad semántica y la comprensión del

contexto, los avances en modelos neuronales y técnicas de aprendizaje automático están mejorando significativamente las capacidades de los sistemas de PLN en este ámbito.

Referencias

- Bender, E. (2011). On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology*, 6(3), 1–26. <https://doi.org/10.33011/lilt.v6i.1239>
- Bender, E. y Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bescós, A. (2020). *Ambigüedad*. <https://www.archiletras.com/actualidad/la-ambigüedad-nos-rodeares-capaz-de-verla/>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., y Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33 (1), 1877–1901. <https://doi.org/10.48550/arXiv.2005.14165>
- Cambria, E., y White, B. (2014). *Jumping NLP curves: A review of natural language processing research*. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. <https://doi.org/10.1109/MCI.2014.2307227>
- Carnie, A. (2021). *Syntax: A generative introduction (4th ed.)*. Wiley-Blackwell. <https://www.wiley.com/en-us/Syntax%3A+A+Generative+Introduction%2C+4th+Edition-p-9781119569312>
- Chipana, L. (2014). *Clases de Morfemas*. <https://serviciodelenguajeyliteratura.blogspot.com/2014/04/morfologia-la-estructura-de-la-palabra.html>
- Codificando Bits. (2025). *¿Qué es una Red Neuronal?* <https://codificandobits.com/blog/que-es-una-red-neuronal/>
- DALL-E. (2025). *Vi al hombre con el telescopio*. <https://chatgpt.com/>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., ... & Welty, C. (2010). *Building Watson: An overview of the DeepQA project*. *AI Magazine*, 31(3), 59–79. <https://doi.org/10.1609/aimag.v31i3.2303>
- Fishwick, P. (2013). *Classic Turing Machine with Tape Erasure*. <https://medium.com/creative-automata/classic-turing-machine-with-tape-erasure-e14870ad154e>
- Gomez-Uribe, C. A., & Hunt, N. (2016). *The Netflix recommender system: Algorithms, business value, and innovation*. *ACM Transactions on Management Information Systems*, 6(4), 1–19. <https://doi.org/10.1145/2843948>
- Honnibal, M., y Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*. 7(1), 411–420. <https://spacy.io/>

- Hutchins, W. J. (2005). *The history of machine translation in a nutshell*. Machine Translation, 19(1), 1–21. <https://aclanthology.org/www.mt-archive.info/10/Hutchins-2014.pdf>
- Jurafsky, D., y Martin, J. H. (2021). *Speech and language processing (3rd ed.)*. Pearson. https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. <https://aclanthology.org/J16-3008.pdf>
- Manning, C. D., y Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press. https://icog-labs.com/wp-content/uploads/2014/07/Christopher_D._Manning_Hinrich_Sch%C3%BCtze_Foundations_Of_Statistical_Natural_Language_Processing.pdf
- MedImaging. (2017). *IBM presenta solución Watson Health para imagenología cognitiva*. <https://www.medimaging.es/imagenologia-general/articulos/294768738/ibm-presenta-solucion-watson-health-para-imagenologia-cognitiva.html>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>
- Moreno, A. (2000). *Árbol Sintáctico*. <http://elies.rediris.es/elies9/3-1-2.htm>
- Navigli, R. (2009). *Word sense disambiguation: A survey*. ACM Computing Surveys, 41(2), 1–69. <https://doi.org/10.1145/1459352.1459355>
- Palmer, M., Gildea, D., & Xue, N. (2010). *Semantic role labeling*. Synthesis Lectures on Human Language Technologies, 3(1), 1–103. <https://web.stanford.edu/~jurafsky/slp3/21.pdf>
- Peláez, B. y Storm, Q. (2022) *Tokenización*. <https://www.getapp.es/blog/2403/que-es-pln-procesamiento-de-lenguaje-natural>
- Relinns Technologies. (2025). *What is N-Gram and How does it work?* <https://botpenguin.com/glossary/n-gram>
- Rodríguez, M. y Sánchez F. (2018). *¿Qué estudia la morfología?* <https://albordeodelengua.blogspot.com/2018/10/que-estudia-la-morfologia-esquema-para.html>
- Susskind, R. (2017). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press. Susskind, R. (2017). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.
- Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books. <https://psnet.ahrq.gov/issue/deep-medicine-how-artificial-intelligence-can-make-healthcare-human-again>
- Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433–460. Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433–460.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems, 30, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. arXiv preprint arXiv:1609.08144. <https://doi.org/10.48550/arXiv.1609.08144>

Términos para definición

Lingüística

La lingüística es la ciencia que estudia el lenguaje en todas sus manifestaciones, incluyendo su estructura, evolución y función en la comunicación humana. Se divide en diversas ramas, como la fonética y fonología, que analizan los sonidos del lenguaje; la morfología, que estudia la estructura de las palabras; la sintaxis, encargada de las reglas para formar oraciones; la semántica, que investiga el significado; y la pragmática, que examina el uso del lenguaje en contextos específicos.

Además, la lingüística abarca especialidades como la sociolingüística, que estudia la relación entre el lenguaje y la sociedad; la psicolingüística, que investiga los procesos cognitivos del lenguaje; y la lingüística computacional, aplicada al procesamiento del lenguaje por máquinas. Su estudio es esencial para comprender la evolución de las lenguas, mejorar la comunicación, preservar idiomas en peligro y desarrollar tecnologías como la traducción automática y la inteligencia artificial.

Modelos de lenguaje

Los modelos de lenguaje son sistemas computacionales diseñados para comprender, generar y procesar texto en lenguaje natural. Estos modelos se basan en técnicas de inteligencia artificial y aprendizaje automático para predecir la probabilidad de ocurrencia de palabras o secuencias de palabras en un contexto determinado.

En términos generales, un modelo de lenguaje aprende patrones lingüísticos a partir de grandes volúmenes de datos textuales. Utiliza estos patrones para completar frases, traducir textos, responder preguntas y generar contenido de manera coherente.



La excelencia no se improvisa

síguenos

