



Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

# Procesamiento de Lenguaje Natural: ¿Cómo entienden las computadoras el lenguaje humano?

Mauricio Toledo-Acosta  
[mauricio.toledo@unison.mx](mailto:mauricio.toledo@unison.mx)

Departamento de Matemáticas  
Universidad de Sonora



Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## Section 1

# ¿Qué es el PLN?



# Procesamiento de Lenguaje Natural

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

El procesamiento de lenguaje natural (PLN) es un campo de las ciencias de la computación que estudia las interacciones entre las computadoras y el lenguaje humano, así como los detalles computacionales de las lenguas naturales.

El PLN combina modelos de lingüística computacional y el aprendizaje automáticos para procesar el lenguaje humano



# Distintivos Niveles de Abstracción

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

El análisis del NLP puede ocurrir a distintos niveles cognitivos:

- Nivel de caracteres.
- Nivel de tokens/palabras.
- Nivel de frases.
- Nivel de oraciones.
- Nivel de documentos.

Además, podemos ver la sintaxis y semántica:

- Incoloras ideas verdes duermen furiosamente.
- Mi niñas duerme tranquilo.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- **Análisis de sentimientos.**
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.





# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



# Algunas tareas del PLN

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Las tareas pueden decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependencia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## Section 2

# El PLN en la actualidad





# Ejemplos de PLN: Llenado de formas

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

### Name form

\* Required

Email \*

Cannot pre-fill email

Name

Your answer

Get link

### Contacts

Name

Phone Area

Address

City

State Zip

Email

Birthday



# Ejemplos de PLN: Llenado de formas

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Name form

\* Required

Email \*

Cannot pre-fill email

Name

Your answer

Get link

Contacts

Name

Phone

Area

Address

City

State

Zip

Email

Birthday

Retos: Mayúsculas/mínusculas, caracteres invalidos, formatos.



# Ejemplos de PLN: Búsquedas

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía



Ejemplos de búsqueda: Búsqueda 1, Búsqueda 2, Búsqueda 3



# Ejemplos de PLN: Búsquedas

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía



Ejemplos de búsqueda: Búsqueda 1, Búsqueda 2, Búsqueda 3

Retos: Mayúsculas/mínusculas, sintaxis, tipos, caracteres  
invalidos, semántica etc.

BERT en las búsquedas



# Ejemplos de PLN: Large Language Models

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Los LLMs son modelos de lenguaje diseñados para procesar y generar texto de manera autónoma. Es capaz de mantener conversaciones con usuarios en lenguaje natural, responder preguntas, proporcionar información, traducción automática, resumen de texto, todo esto a través de texto. Su versatilidad radica en su capacidad para comprender y generar texto coherente y relevante en una amplia gama de contextos y aplicaciones.

Detección de texto generado por AI



# Un poco más sobre ChatGPT 3

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Ha obtenido los siguientes puntajes:

- Obtuvo 1020 puntos de 1600 posibles en los exámenes SAT.
- Obtuvo un nivel de C+ en los exámenes de cuatro cursos de la Facultad de Derecho de la Universidad de Minnesota.
- Obtuvo un nivel de B a B- en el examen de un curso de gestión empresarial de la Wharton School of Business de la Universidad de Pensilvania.
- Obtuvo una puntuación igual o cercana al aprobado en el Examen de Licencia Médica de EE.UU., que los médicos deben aprobar para obtener la licencia médica.



# Las alucinaciones

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

ChatGPT y otros LLMs para generación de textos no son infalibles.

- Acerca de si los dinosaurios construyeron una civilización:  
*Some species of dinosaurs even developed primitive forms of art, such as engravings on stones.* Corrección
- When did Leonardo da Vinci paint the Mona Lisa?  
*Leonardo da Vinci painted the Mona Lisa in 1815.* Corrección
- Ha inventado referencias y documentos.
- Puede reproducir sesgos.
- Casos concretos Gemini, Claude.



# Tendencias en la investigación

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

Ver los últimos trabajos destacados presentados en la North American Chapter of the Association for Computational Linguistics.

<https://2024.naacl.org/>

¿Qué temáticas y tendencias observas?





Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## Section 3

# El lenguaje natural



# El lenguaje natural

## Procesamiento de Lenguaje Natural:

### ¿Qué es el PLN?

### El PLN en la actualidad

### El lenguaje natural

### Temario

### Bibliografía

Un **lenguaje natural** es cualquier lenguaje que se produce de forma natural en una comunidad humana mediante un proceso de uso, repetición y cambio. Puede adoptar diferentes formas, normalmente una lengua hablada o una lengua de signos. Las lenguas naturales se distinguen de las lenguas construidas y formales, como las que se utilizan para programar ordenadores o estudiar la lógica.



# Los dificultades del lenguaje natural

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- **Ambigüedad.** Una misma palabra o frase puede tener múltiples significados dependiendo del contexto. Ejemplos: Banco, vela, planta.
- **Variabilidad lingüística.** Existen muchas variaciones en la forma en que las personas utilizan el lenguaje, incluyendo diferencias regionales, culturales y generacionales. Ejemplos: Padre, chido, cool, genial, etc.
- **Ironía y sarcasmo.** Son difíciles de detectar debido a su naturaleza no literal.



# Los dificultades del lenguaje natural

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- **Resolución de la correferencia.** Identificar a qué se refieren pronombres como “él”, “ella”, “eso” en un texto. Se requiere comprensión del discurso y conocimiento del mundo. Ejemplo: María le dio un libro a Ana. Ella lo leyó rápidamente.
- **Ambigüedades.** ¿Cuál es el significado de las siguientes frases?
  - Vi al niño con los binoculares.
  - I shot my neighbor in my pijamas.



# Los dificultades del lenguaje natural

## Procesamiento de Lenguaje Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- **Procesamiento de sentimientos y emociones.** Capturar el tono, el sentimiento y las emociones detrás del lenguaje es un desafío, ya que requiere comprensión de las sutilezas del discurso humano.
- **Tratamiento de errores y ruido.** Los errores gramaticales, los errores tipográficos y otros tipos de ruido en el texto pueden dificultar el procesamiento del lenguaje natural.



# ¿Cuál es la solución?

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

La solución puede ser más o menos sofisticada de acuerdo a la tarea que se quiera resolver.

Puede requerir un algoritmo muy sencillo o un modelo tan avanzado como un Large Language Model (LLM), por ejemplo ChatGPT, LLaMa, DeepSeek, Qwen, etc.



Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## Section 4

# Temario



# Temario

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- 1 Introducción:
  - Tareas del NLP
  - Expresiones regulares
- 2 Métodos de modelado clásico de NLP
  - N-gramas.
  - Modelos BOW (1954), TF-IDF (1970s).
- 3 Introducción al Machine Learning
- 4 Algoritmos Clásicos del NLP
  - Naive Bayes (late 1700s - )
  - Latent Dirichlet Allocation (2000)
  - Latent Semantic Analysis (1988)





# Temario

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## 5 Métodos de redes neuronales de NLP

- Embeddings de tokens y documentos.
- Redes recurrentes y LSTM.
- Modelos Word2Vec (Google, 2013), Glove (Stanford, 2014), FastText (FB, 2016), Doc2Vec (Google, 2014), WordRank (2016).

## 6 Transformadores y LLM.

- Arquitectura Transformers (2017): *Attention is all you need*.
- BERT (Google, 2018), RoBERTa (FB, 2019), T5 (2020), GTP-3 (OpenAI, 2020), LLaMA (Meta AI, 2022), GPT-4 (Open AI, 2023).
- Afinación de modelos pre-entrenados.
- Prompt Engineering.
- Retrieval Augmented Generation (RAG).



# Tareas a realizar

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- Limpieza de texto y preprocesamiento.
- Análisis de Sentimientos y Clasificación.
- Modelaje de tópicos.
- Information Retrieval.



# Evaluación

Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

La evaluación del curso será mediante la elaboración y presentación de proyectos y actividades (teóricos y/o prácticos):

- Presentaciones en equipo (2x)
- Proyectos en equipo (2x)
- Proyectos individuales (2x)
- Actividades en clase (1x)
- Tareas individuales (1x)
- Exámenes escritos (2x)

Al final la calificación será un promedio con estos pesos.



Procesamiento  
de Lenguaje  
Natural:

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

## Section 5

# Bibliografía



# Bibliografía

Procesamiento  
de Lenguaje  
Natural

¿Qué es el  
PLN?

El PLN en la  
actualidad

El lenguaje  
natural

Temario

Bibliografía

- Goldberg, Y. (2022). **Neural network methods for natural language processing**. Springer Nature.
- Eisenstein, J. (2018). **Natural language processing**.
- Jurafsky, D., Martin, J. H. (2019). **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**.