

ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL

Procesamiento de Lenguaje Natural (NPL)



TEMARIO:

Semana 1 – Estructura del lenguaje natural

Semana 2 – Procesamiento léxico y morfológico

Semana 3 – Librerías para el NPL

Semana 4 – Aplicar redes neuronales al NPL

Semana 5 – Aplicación del NPL

Semana 6 – Programación de la inteligencia artificial

Semana 7 – Uso de las redes neuronales recurrentes

Semana 8 – Usos de GPT

METODOLOGÍAS ACTIVAS

Estudios de caso:

- Calentamiento sin código (conceptos clave)
- Tareas del PLN
- Enfoques

Definición de Procesamiento de Lenguaje Natural (PLN / NPL)

El **Procesamiento de Lenguaje Natural (PLN)** es un área de la inteligencia artificial que busca que las computadoras **entiendan, interpreten y generen lenguaje humano** (escrito o hablado).

Se nutre de varias disciplinas:

Lingüística → Reglas gramaticales, estructura del lenguaje, significados.

Informática → Algoritmos, programación, estructuras de datos.

Psicología → Cómo los humanos procesan y comprenden el lenguaje.

Ingeniería → Creación de sistemas aplicados (chatbots, traductores, buscadores)

Analizar los problemas del PLN significa detectar las dificultades que surgen cuando intentamos que la máquina entienda lo que para los humanos es natural (ambigüedades, ironía, sinónimos, contexto, etc.).

Estructura del lenguaje natural en PLN (Procesamiento de Lenguaje Natural)

En **Procesamiento de Lenguaje Natural (PLN)**, cuando hablamos de la **estructura del lenguaje natural**, nos referimos a cómo se organiza y analiza el lenguaje humano para que una máquina pueda procesarlo.

Esta estructura se adapta desde la lingüística, pero con un enfoque computacional.
Se suele dividir en **múltiples niveles**

Nivel	Qué analiza	Ejemplo
Fonético / Fonológico	Sonidos, pronunciación, entonación (aplica a voz).	“casa” vs “caza” → suenan parecido, significan distinto.
Morfológico	La forma de las palabras: raíces, afijos, conjugaciones.	estudiantes → raíz: estudiar + sufijo -antes.
Léxico	Palabras como unidades de un diccionario mental.	“banco” → puede ser institución financiera o asiento.
Sintáctico	Reglas gramaticales y estructura de oraciones.	“El perro muerde al gato” ≠ “El gato muerde al perro”.
Semántico	El significado de palabras y oraciones.	“Python” → ¿lenguaje de programación o serpiente?
Pragmático	El uso del lenguaje en contexto real.	“¿Puedes abrir la ventana?” → no pregunta por capacidad, es una petición.
Discursivo	Cómo se conectan varias oraciones en un texto coherente.	En una noticia: introducción, desarrollo, conclusión.

Enfoques en PLN

Simbólico (basado en reglas)

Usa gramáticas, reglas lingüísticas y diccionarios.

Ejemplo: “Si la palabra termina en *-ar*, probablemente es un verbo en infinitivo.”

Problema: requiere miles de reglas → difícil de mantener.

Estadístico (basado en datos)

Usa probabilidades y frecuencia de palabras en grandes corpus.

Ejemplo: *bag of words* → contar cuántas veces aparece cada palabra.

Problema: pierde el orden y el contexto.

Hoy en día se combinan con redes neuronales y modelos de deep learning (lo veremos más adelante).

Concepto de *Corpus* y Tokenización

Corpus

Conjunto grande de **textos** (ej: Wikipedia, noticias, tuits) que se usan para entrenar modelos de lenguaje.

Tokenización

Dividir un texto en piezas manejables (palabras, frases, o incluso sub-palabras).

Ejemplo:

Texto = “*Los modelos de lenguaje son poderosos*”

Tokens: ["Los", "modelos", "de", "lenguaje", "son", "poderosos"]

La Metáfora

Enseñar PLN es como enseñar a un extranjero a cocinar en nuestro país:

Lingüística → Le enseñas la receta y el idioma de los ingredientes.

Informática → Le das utensilios y técnicas de cocina (algoritmos).

Psicología → Le explicas *cómo* los humanos deciden qué plato cocinar.

Ingeniería → Lo llevas a una cocina real con clientes (aplicación práctica).

- **El corpus** sería su *libro de recetas* (miles de ejemplos).
- **La tokenización** serían los *ingredientes* (tomates, cebollas, arroz).
- **Los enfoques simbólico y estadístico** son como dos formas de aprender:

Memorizar reglas (simbólico).

Probar y contar cuántas veces un plato sale rico (estadístico).

Estudio de caso 1

Texto:

“Los estudiantes de ingeniería usan Python para proyectos de inteligencia artificial en la universidad.”

Fases de preprocesamiento:

1.Segmentación: Detectar oraciones.

→ Aquí hay solo una.

2.Tokenización: ["Los", "estudiantes", "de", "ingeniería", "usan", "Python", "para", "proyectos", "de", "inteligencia", "artificial", "en", "la", "universidad"]

3.Normalización: pasar a minúsculas → "python"

4.Limpieza: quitar puntuación/stopwords → ["estudiantes", "ingeniería", "usan", "python", "proyectos", "inteligencia", "artificial", "universidad"]

5.Lematización: "usan" → "usar"

6.Vectorización (estadístico): contar frecuencia de cada palabra.

Estudio de caso 2

Texto:

"Los algoritmos modernos de aprendizaje automático permiten clasificar grandes volúmenes de datos."

Tokenización: ["Los", "algoritmos", "modernos", "de", "aprendizaje", "automático", "permiten", "clasificar", "grandes", "volúmenes", "de", "datos"]

Normalización: ["los", "algoritmos", "modernos", "de", "aprendizaje", "automático", "permiten", "clasificar", "grandes", "volúmenes", "de", "datos"]

Limpieza (sin stopwords): ["algoritmos", "modernos", "aprendizaje", "automático", "permiten", "clasificar", "grandes", "volúmenes", "datos"]

Lematización: ["algoritmo", "moderno", "aprendizaje", "automático", "permitir", "clasificar", "grande", "volumen", "dato"]

Problemas detectados:

- 1."automático" puede ser adjetivo o sustantivo según contexto.
- 2."volúmenes" → ¿se mantiene como "volumen" o se pierde el matiz de pluralidad?
- 3."aprendizaje automático" es un concepto técnico, no debería dividirse.

Estudio de caso 3

Texto:

"El presidente anunció nuevas medidas económicas para apoyar a las pequeñas empresas."

Tokenización: ["El", "presidente", "anunció", "nuevas", "medidas", "económicas", "para", "apoyar", "a", "las", "pequeñas", "empresas"]

Normalización: ["el", "presidente", "anunció", "nuevas", "medidas", "económicas", "para", "apoyar", "a", "las", "pequeñas", "empresas"]

Limpieza: ["presidente", "anunció", "nuevas", "medidas", "económicas", "apoyar", "pequeñas", "empresas"]

Lematización: ["presidente", "anunciar", "nuevo", "medida", "económico", "apoyar", "pequeña", "empresa"]

Problemas detectados:

"presidente" puede referirse a un cargo o a una persona concreta.

"medidas económicas" debe verse como unidad semántica.

"nuevas" → se pierde el matiz de plural/femenino al pasar a lema.

Estudio de caso 4

Texto:

"Los estudiantes de ingeniería usan Python para proyectos de inteligencia artificial en la universidad."

Tokenización: ["Los", "estudiantes", "de", "ingeniería", "usan", "Python", "para", "proyectos", "de", "inteligencia", "artificial", "en", "la", "universidad"]

Normalización: ["los", "estudiantes", "de", "ingeniería", "usan", "python", "para", "proyectos", "de", "inteligencia", "artificial", "en", "la", "universidad"]

Limpieza: ["estudiantes", "ingeniería", "usan", "python", "proyectos", "inteligencia", "artificial", "universidad"]

Lematización: ["estudiante", "ingeniería", "usar", "Python", "proyecto", "inteligencia artificial", "universidad"]

Problemas detectados:

"Python" no debe traducirse a "pitón".

"inteligencia artificial" debe tratarse como concepto compuesto.

"usan" puede perder información temporal (tiempo verbal).

Pautas de aprendizaje específico

Analizar problemas del PLN significa:

- i. **Detectar ambigüedad semántica** (una palabra con varios significados).
- ii. **Identificar dependencia del contexto** (qué significa “ellos” depende de la oración anterior).
- iii. **Notar variabilidad morfológica** (correr, corría, corrimos).
- iv. **Reconocer ruido en el texto** (errores ortográficos, emojis, hashtags).

Pautas para Lematizar

La **lematización** consiste en reducir una palabra a su forma base.

- **Identifica el verbo en infinitivo** → "usando" → *usar*, "corría" → *correr*.
- **Identifica el sustantivo en singular** → "estudiantes" → *estudiante*.
- **Adjetivos en su forma básica** → "mejores" → *mejor*.
- **Mantén nombres propios y acrónimos** → "Python", "ONU".
- **Cuida expresiones compuestas** → "inteligencia artificial" debe tratarse como unidad.

Actividad práctica

Objetivo: Identificar problemas y fases de preprocesamiento.

Tomar uno de los 3 casos de análisis y responder el padlet

Realizar:

1. **Tokenización:** separa palabras.
2. **Normalización:** convierte a minúsculas.
3. **Limpieza:** elimina stopwords y signos de puntuación.
4. **Lematización:** intenta reducir palabras a su raíz
(mientras vemos las librerías de python, hazlo intuitivamente).

Escribe 3 problemas detectados en el texto:

1. **Palabras polisémicas** (“banco” puede ser institución o asiento).
2. Nombres propios (“Python” no debería convertirse en “pitón”).
3. Expresiones compuestas (“inteligencia artificial” debe tratarse como un concepto, no dos palabras separadas).

Actividad práctica – casos de análisis

Texto 1 – Salud

“Los médicos recomiendan realizar actividad física tres veces por semana para mejorar la salud cardiovascular.”

Texto 2 – Economía

“El mercado financiero mostró una caída significativa debido a la inflación y la inestabilidad política en la región.”

Texto 3 – Tecnología

“La inteligencia artificial está revolucionando la forma en que las empresas gestionan datos y procesos de negocio.”

<https://padlet.com/sergiopuertomo/tareas-del-pln-5hxgrlhxg5tmx8yd>

ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL

Procesamiento de Lenguaje Natural (NPL)



TEMARIO:

Semana 1 – Estructura del lenguaje natural

Semana 2 – Procesamiento léxico y morfológico

Semana 3 – Librerías para el NPL

Semana 4 – Aplicar redes neuronales al NPL

Semana 5 – Aplicación del NPL

Semana 6 – Programación de la inteligencia artificial

Semana 7 – Uso de las redes neuronales recurrentes

Semana 8 – Usos de GPT



Su dirección de correo electrónico se compartirá con el propietario de la encuesta.

1. ¿Cómo calificarías tu experiencia en la sesión que acabas de tomar? ★★★★★ *

- ☐ ★ Nada satisfecho
- ☐ ★★ Poco satisfecho
- ☐ ★★★ Neutral
- ☐ ★★★★ Muy satisfecho
- ☐ ★★★★★ Totalmente satisfecho

Al responder a esta encuesta, acepta la [Declaración de privacidad](#) y los [Términos del servicio](#) de Zoom.

Enviar

METODOLOGÍAS ACTIVAS

Estudios de caso:

- Aplicación para: *Estructura del lenguaje natural – Colab*
- *Casos de estudio de la semana*

Procesamiento léxico y morfológico en PLN

Léxico = qué es la palabra y qué significa.

Morfológico = cómo cambia la palabra según reglas gramaticales.

Procesamiento léxico y morfológico en PLN

PROCESAMIENTO LÉXICO

Se refiere al **análisis de las palabras como unidades mínimas del lenguaje**.

En esta etapa se busca:

- **Identificar palabras** en un texto (**tokenización**).
- **Asignar categorías gramaticales** (**part-of-speech tagging, POS**).
- **Acceder a información semántica:** **significado, sinónimos, antónimos**.
- **Reconocer expresiones compuestas** (**inteligencia artificial, red neuronal**).

*En PLN, el léxico funciona como un **diccionario computacional** que permite saber qué es cada palabra y cómo se comporta en una oración.*

Procesamiento léxico y morfológico en PLN

PROCESAMIENTO MORFOLÓGICO

Se centra en la **estructura interna de las palabras** y cómo estas cambian según reglas de la lengua. Incluye:

Flexión: variaciones de género, número, tiempo, modo, aspecto.

Ejemplo: *correr* → *corría*, *corrieron*.

Derivación: creación de nuevas palabras con sufijos/prefijos.

Ejemplo: *feliz* → *infeliz*, *felicidad*.

Composición: unión de dos palabras.

Ejemplo: *paraguas*, *baloncesto*.

En PLN, esto es esencial para:

- Reducir palabras a su **forma base** (**lematización**).
- Evitar confusión entre variaciones superficiales (**mejores**, **mejor**).
- Detectar relaciones semánticas (**correr** ↔ **corriendo** ↔ **corrí**).

Procesamiento léxico y morfológico en PLN

Importancia en PLN

- ✓ **Ambigüedad gramatical:** una palabra puede pertenecer a varias categorías. Ejemplo: *banco* (institución / asiento).
- ✓ **Normalización del texto:** trabajar con palabras base mejora modelos estadísticos y neuronales.
- ✓ **Comprensión semántica:** permite que la máquina capte relaciones entre palabras (*médico* ↔ *medicina*).
- ✓ **Aplicaciones:** correctores ortográficos, traductores automáticos, buscadores semánticos, chatbots.

Procesamiento léxico y morfológico en PLN

Ejemplo conceptual

Texto: “*Los médicos recomiendan realizar actividad física tres veces por semana.*”

Léxico:

Identificar categorías: *médicos* (sustantivo plural), *recomiendan* (verbo en presente).

Morfológico:

médicos → **base:** *médico*, plural masculino.

recomiendan → **verbo** *recomendar*, tiempo presente, tercera persona plural.

El **procesamiento léxico** actúa como el **diccionario**,
y el **procesamiento morfológico** como la **gramática** que explica cómo cambian las palabras.

Tabla comparativa – Procesamiento léxico vs. morfológico

Aspecto	Procesamiento léxico	Procesamiento morfológico
Definición	Analiza las palabras como unidades del lenguaje (diccionario computacional).	Analiza la estructura interna de las palabras y sus variaciones.
Enfoque	Identificación de palabras y sus categorías gramaticales.	Estudio de prefijos, sufijos, raíces, género, número, tiempo verbal.
Tareas principales	<ul style="list-style-type: none">- Tokenización.- Etiquetado gramatical (POS tagging).- Detección de expresiones compuestas.- Relación con significados.	<ul style="list-style-type: none">- Flexión (género, número, tiempo).- Derivación (nuevas palabras con prefijos/sufijos).- Composición (unión de palabras).
Ejemplo	médicos → sustantivo plural, masculino.recomiendan → verbo presente, 3ª persona plural.	médicos → base: médico + sufijo plural “-s”.recomiendan → recomendar + desinencia verbal “-an”.
Resultado esperado	Saber qué rol cumple cada palabra en la oración.	Obtener la forma base (lema) y entender variaciones morfológicas.
Aplicaciones	<ul style="list-style-type: none">- Análisis semántico.- Chatbots.- Motores de búsqueda.- Traducción automática.	<ul style="list-style-type: none">- Correctores ortográficos.- Lematización.- Generación automática de lenguaje.- Detección de relaciones semánticas.

Las metáforas

Metáfora	Léxico (identidad de la palabra)	Morfológico (cambios de la palabra)
El carro y sus marchas 🚗	El carro es siempre el mismo (Toyota Corolla).	Las marchas cambian cómo se mueve: 1ra, 2da, reversa.
La receta y los ingredientes 🍲	La receta es fija: “arroz con pollo”.	Los ingredientes se modifican: arroz integral, pollo asado, arroz blanco.
La canción y los covers 🎵	La canción original siempre es la misma (“Imagine” de John Lennon).	Los covers cambian: en rock, salsa, acústico.
El árbol y sus estaciones 🌳	El árbol es el mismo (roble).	En invierno sin hojas, en primavera florecido, en otoño con hojas secas.
La película y los doblajes 🎬	La película es la misma (“Toy Story”).	Los doblajes cambian: español, inglés, francés.

El **léxico** es **constante**: carro, receta, canción, árbol, película.

El **morfológico** son las **variaciones** de esa base: marchas, ingredientes, covers, estaciones, doblajes.

Los ejemplos

1. Ejemplo: *cantar, cantaba, cantó*

Léxico: la palabra base es “**cantar**” → un **verbo**.

Morfológico: se transforma según tiempo verbal: *cantaba* (pretérito imperfecto), *cantó* (pretérito perfecto).

Esto muestra cómo una sola palabra léxica tiene muchas formas morfológicas.

2. Ejemplo: *feliz, infeliz, felicidad*

Léxico: la palabra base es “**feliz**” → un **adjetivo**.

Morfológico: a partir de la raíz surgen variaciones:

- *infeliz* (prefijo → negación).
- *felicidad* (sufijo → convierte adjetivo en sustantivo).

Aquí se ve la derivación morfológica, cómo la misma raíz genera distintas palabras relacionadas.

Los ejemplos

3. Ejemplo: *niño, niños, niña, niñas*

Léxico: la palabra base es “**niño**” → un **sustantivo**.

Morfológico: cambia con género y número:

masculino/femenino → *niño / niña*.

singular/plural → *niño / niños, niña / niñas*.

*Aquí vemos la **flexión morfológica** (cambios por reglas gramaticales).*

4. Ejemplo: *correr, corrí, corriendo*

Léxico: la palabra base es “**correr**” → un **verbo**.

Morfológico: variaciones según persona y tiempo:

corrí (1ª persona, pasado).

corriendo (gerundio).

*Aquí se ve cómo la **forma verbal cambia**, pero el léxico (la acción de correr) se mantiene.*

Los ejemplos

5. Ejemplo: *médico, médicos, médica, médicas*

Léxico: la palabra base es “**médico**” → un **sustantivo**.

Morfológico: cambia con género y número:

médico / médica (género).

médico / médicos, médica / médicas (número).

*Otra vez vemos la **flexión morfológica** aplicada a sustantivos.*

Léxico =

el significado principal, la raíz o palabra base (ejemplo: *cantar, feliz, niño, correr, médico*).

Morfológico =

todas las variaciones que esa palabra puede tener (tiempo verbal, género, número, sufijos, prefijos).

Tabla – Léxico vs. Morfológico / ejemplos

Ejemplo	Procesamiento léxico (palabra base / identidad)	Procesamiento morfológico (variaciones de la palabra)
cantar, cantaba, cantó	cantar → verbo en infinitivo	cantaba (pretérito imperfecto), cantó (pretérito perfecto)
feliz, infeliz, felicidad	feliz → adjetivo	infeliz (con prefijo), felicidad (con sufijo, sustantivo)
niño, niños, niña, niñas	niño → sustantivo	Cambios de género y número: niños, niña, niñas
correr, corrí, corriendo	correr → verbo en infinitivo	corrí (pasado), corriendo (gerundio)
médico, médicos, médica, médicas	médico → sustantivo	Variaciones de género y número: médicos, médica, médicas

Léxico = responde a “¿qué es esta palabra?” (su identidad fija).

Morfológico = responde a “¿cómo puede cambiar esta palabra?” (tiempo, número, género, prefijos, sufijos).