

# Syllabus de Procesamiento de Lenguaje Natural Avanzado

---

## Materia: Procesamiento de Lenguaje Natural Avanzado

<https://www.iimas.unam.mx/wp-content/uploads/2024/11/Asignaturas.pdf>

### OBJETIVOS DEL CURSO

#### General:

Analizar, diseñar e implementar soluciones modernas de procesamiento del lenguaje natural basadas en modelos de lenguaje a gran escala, técnicas de prompting, recuperación aumentada con generación y agentes, partiendo de una revisión sintética de los fundamentos y enfatizando la evaluación crítica, las limitaciones y los aspectos éticos de estos enfoques.

#### Específicos:

- Consolidar de manera sintética los fundamentos del procesamiento del lenguaje natural necesarios para el estudio de enfoques modernos.
- Analizar la transición de representaciones clásicas del lenguaje hacia representaciones contextualizadas y modelos de lenguaje a gran escala.
- Diseñar y aplicar técnicas modernas de PLN, incluyendo prompting, recuperación aumentada con generación y agentes basados en modelos de lenguaje.
- Evaluar de forma crítica sistemas de PLN modernos mediante métricas, análisis de errores y consideraciones éticas.
- Desarrollar un proyecto aplicado o de investigación que integre de manera coherente técnicas avanzadas de procesamiento del lenguaje natural.

### CONTENIDO TEMÁTICO DEL CURSO

#### 1. Revisión de fundamentos de PLN

- 1.1. Revisión sintética de PLN clásico
- 1.2. Limitaciones de embeddings estáticos
- 1.3. Introducción a representaciones contextualizadas

#### 2. Representaciones contextualizadas del lenguaje, Transformers y modelos preentrenados

- 2.1. Contextual embeddings
- 2.2. Modelos preentrenados (BERT, RoBERTa, GPT, T5)
- 2.3. Fine-tuning vs feature extraction
- 2.4. Análisis semántico contextual
- 2.5. Sesgos y limitaciones

#### 3. Modelos de lenguaje a gran escala (LLMs)

- 3.1. Evolución de los modelos de lenguaje
- 3.2. Arquitectura Transformer (visión funcional)
- 3.3. Capacidades emergentes de los LLMs
- 3.4. Limitaciones, alucinaciones y costos

#### 4. Prompting y aprendizaje basado en instrucciones

- 4.1. Zero-shot y few-shot prompting
- 4.2. Prompt engineering
- 4.3. Chain-of-Thought y variantes
- 4.4. Evaluación de prompts

#### 5. Recuperación aumentada con generación (RAG)

- 5.1. Motivación: límites del conocimiento paramétrico
- 5.2. Arquitectura RAG
- 5.3. Indexación, embeddings y retrieval
- 5.4. Evaluación de sistemas RAG

# Syllabus de Procesamiento de Lenguaje Natural Avanzado

## 6. Agentes basados en modelos de lenguaje

- 6.1. Concepto de agentes LLM
- 6.2. Planeación, memoria y herramientas
- 6.3. Orquestación de agentes
- 6.4. Casos de uso y limitaciones

## 7. Evaluación, ética y aplicaciones avanzadas

- 7.1. Evaluación de sistemas basados en LLMs
- 7.2. Análisis de errores y confiabilidad
- 7.3. Aspectos éticos y sociales
- 7.4. Aplicaciones avanzadas y tendencias actuales

### PROGRAMA GENERAL DE ACTIVIDADES DEL CURSO

A continuación, se describen las actividades que se desarrollarán durante el semestre en la materia de Procesamiento de Lenguaje Natural, la cual será impartida por la Dra. Helena Gómez Adorno y el Dr. Fazlourrahman Balouchzahi.

Semana	Profe	Fechas	Temas
1	Helena	3 y 5 feb	<b>Teoría:</b> Introducción general del curso: reglas, evaluación y proyecto. Revisión sintética de los fundamentos del PLN.
2	Helena	10 y 12 feb	<b>Teoría:</b> Limitaciones de embeddings estáticos y motivación de representaciones contextualizadas. <b>Práctica:</b> Uso de embeddings contextualizados (BERT-like) como extractores de características. <b>Tema:</b> Representaciones contextualizadas del lenguaje. <b>Tarea:</b> Comparación Word2Vec vs embeddings contextualizados en similitud semántica.
3	Fazl	17 y 19 feb	<b>Teoría:</b> Fine-tuning vs feature extraction. Sesgos y limitaciones de representaciones contextualizadas. <b>Práctica:</b> Fine-tuning ligero para una tarea de clasificación/etiquetado (dataset pequeño). <b>Tema:</b> Representaciones contextualizadas: ajuste y análisis. <b>Tarea:</b> Reporte breve de resultados + análisis de errores.
4	Helena	24 y 26 feb	<b>Teoría:</b> Transformer (visión funcional) aplicado a LLMs: atención, contexto y escalamiento. <b>Práctica:</b> Uso de un modelo preentrenado en tareas tipo: clasificación, QA o resumen (pipeline). <b>Tema:</b> Transformers y modelos preentrenados.

## Syllabus de Procesamiento de Lenguaje Natural Avanzado

			<b>Tarea:</b> Experimento reproducible con un modelo preentrenado + discusión de limitaciones.
5	Fazl	3 y 5 mar	<p><b>Teoría:</b> Modelos de lenguaje a gran escala (LLMs): evolución, capacidades y límites (alucinaciones, costo, contexto).</p> <p><b>Práctica:</b> Exploración guiada de comportamiento de un LLM (casos controlados).</p> <p><b>Tema:</b> LLMs: fundamentos y comportamiento.</p> <p><b>Tarea:</b> Ensayo técnico corto: “qué puede y qué no puede hacer un LLM” con ejemplos.</p>
6	Helena	10 y 12 mar	<p><b>Teoría:</b> Prompting: zero-shot, few-shot e instruction prompting. Estructura y buenas prácticas.</p> <p><b>Práctica:</b> Diseño y evaluación de prompts para una tarea (p.ej., clasificación semántica/emocional).</p> <p><b>Tema:</b> Prompting e instrucciones.</p> <p><b>Tarea:</b> Portafolio de prompts con métricas y análisis de fallos.</p>
7	Fazl	17 y 19 mar	<p><b>Teoría:</b> Prompting avanzado: Chain-of-Thought (concepto), self-consistency y estrategias de robustez.</p> <p><b>Práctica:</b> Comparación de variantes de prompting en el mismo problema (ablation simple).</p> <p><b>Tema:</b> Prompting avanzado y razonamiento guiado.</p> <p><b>Tarea:</b> Tabla comparativa de prompts + conclusiones (qué funcionó y por qué).</p>
8	Helena	24 y 26 mar	<p><b>Teoría:</b> RAG: motivación, arquitectura y flujo (retrieval → context → generation).</p> <p><b>Práctica:</b> Diseño de pipeline RAG (a nivel conceptual + prototipo base).</p> <p><b>Tema:</b> Recuperación aumentada con generación (RAG).</p> <p><b>Actividad:</b> Presentación y definición del proyecto (elección: LLM / Prompting / RAG / Agents).</p> <p><b>Tarea:</b> Propuesta de proyecto (1–2 páginas) + dataset/fuente + criterios de evaluación.</p>
9	Fazl	7 y 9 abr	<p><b>Teoría:</b> Embeddings para retrieval, indexación y estrategias de búsqueda (top-k, reranking básico).</p>

## Syllabus de Procesamiento de Lenguaje Natural Avanzado

			<p><b>Práctica:</b> Implementación de RAG (indexación + recuperación + generación) con corpus pequeño/mediano.</p> <p><b>Tema:</b> Implementación de sistemas RAG.</p> <p><b>Tarea:</b> Demo funcional + reporte de diseño (decisiones y supuestos).</p> <p><b>Evaluación:</b> Primer examen parcial.</p>
10	Helena	14 y 16 abr	<p><b>Teoría:</b> Evaluación de LLMs/RAG: calidad, factualidad, cobertura, citación, y análisis de errores.</p> <p><b>Práctica:</b> Evaluación del sistema (métricas automáticas + evaluación manual con rúbrica).</p> <p><b>Tema:</b> Evaluación y confiabilidad en LLM/RAG.</p> <p><b>Tarea:</b> Informe de evaluación del sistema (con errores típicos).</p>
11	Fazl	21 y 23 abr	<p><b>Teoría:</b> Agentes con LLM: concepto, herramientas, planificación y memoria (visión general).</p> <p><b>Práctica:</b> Diseño de un flujo agente-herramientas (tool calling / pipeline).</p> <p><b>Tema:</b> Agentes basados en modelos de lenguaje.</p> <p><b>Tarea:</b> Especificación de un agente (objetivo, herramientas, límites, criterios de éxito).</p>
13	Helena	5 y 7 may	<p><b>Teoría:</b> Ética, sesgos, privacidad, seguridad y confiabilidad en sistemas LLM/RAG/Agentes.</p> <p><b>Práctica:</b> Auditoría rápida de sesgos/errores + mitigaciones básicas (prompts, filtros, trazabilidad).</p> <p><b>Tema:</b> Consideraciones éticas y sociales del PLN moderno.</p> <p><b>Tarea:</b> Checklist de riesgos + plan de mitigación para el proyecto.</p>
14	Fazl, Helena	12 y 14 may	<p><b>Práctica:</b> Integración del proyecto final: mejoras (prompting/RAG/agentes) y preparación de resultados.</p> <p><b>Tema:</b> Integración de métodos avanzados en un sistema completo.</p> <p><b>Tarea:</b> Avance de proyecto (resultados preliminares + análisis).</p>
15	Fazl, Helena	19 y 21 may	<p><b>Deployment</b></p> <p><b>Evaluación:</b> Segundo examen parcial.</p>

## Syllabus de Procesamiento de Lenguaje Natural Avanzado

16		26 y 28 may	<b>Actividad:</b> Revisión y asesoría de proyectos finales (orientación a resultados y claridad experimental). <b>Tarea:</b> Versión casi final del reporte + diapositivas borrador.
17		2 y 4 jun	<b>Actividad:</b> Presentación y entrega de proyectos finales.
18		9 y 11 jun	<b>Evaluación:</b> Exámenes finales.

### EVALUACIÓN DEL CURSO

Evaluación de las actividades y el peso relativo de cada grupo de ellas para conformar la calificación final del curso.

Actividad	Porcentaje
Exámenes Parciales	30%
Prácticas OBLIGATORIAS	40%
Proyecto OBLIGATORIO	30%
<b>Total</b>	<b>100%</b>
Examen Final	100%

Por reglamento general de exámenes, se tienen tres oportunidades para acreditar la materia:

- 1.- Presentar TODOS los elementos correspondientes a la evaluación del curso (tabla anterior) en tiempo y forma. Al obtener un promedio mayor o igual a 7.6 se da por acreditado el curso.
- 2.- Presentar el primer examen final y su calificación se pone en actas.
- 3.- Presentar el segundo examen final y su calificación se pone en actas.

Para las calificaciones con decimales .6 sube al siguiente entero. Ej. 5.6 = 6, 6.6=7.

Por reglamento general de exámenes no se puede presentar final para subir de calificación. No se acredita la materia si se obtiene calificación menor o igual a 5.9 en exámenes finales.

### BIBLIOGRAFIA

1. Practical Natural Language Processing. [Vajjala et al., 2020]
2. Speech and Language Processing [Jurafky & Martin]
3. Efficient Estimation of Word Representations in Vector Space [Mikolov et al., 2013]
4. GloVe: Global Vectors for Word Representation [Pennington et al., 2014][Sitio Oficial]
5. Enriching word vectors with subword information [Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T., 2017][Sitio Oficial]
6. Attention Is All You Need [Vaswani et al., 2017]
7. The Illustrated Transformer [Jay Alammar]
8. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Devlin et al., 2019]
9. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks [Lewis et al., 2020]

**PAGINA DE LA MATERIA:** <https://tuaulavirtual.educatic.unam.mx/course/edit.php?id=16134>

# Syllabus de Procesamiento de Lenguaje Natural Avanzado

---

**Contraseña:**

**HORARIO CLASE VIRTUAL ZOOM:**

<https://cuaieed-unam.zoom.us/j/83350463871>

ID de reunión: 833 5046 3871

**PRÁCTICAS:** Todas las prácticas se entregan en la plataforma Moodle de la materia en la fecha establecida para cada una de ellas.

**PROYECTO:**

El proyecto consistirá en el desarrollo de un sistema de PLN que integre conceptos revisados y técnicas avanzadas del curso.

**Opciones de proyecto:**

- Clasificación semántica o emocional
- Similaridad textual
- Modelado de lenguaje
- PLN multilingüe
- Aplicación de PLN a un dominio específico

**Entregables:**

- Reporte técnico (formato LNCS o similar): Realizar un reporte de la resolución del problema en el formato LNCS de Springer que lo pueden hacer tanto en latex como en word. El formato lo pueden descargar aquí:  
<https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines>
- La estructura mínima del reporte será la siguiente: Título, Autor/es, Resumen (Abstract), Introducción, Metodología, Resultados, Conclusiones, Referencias. Si alguno considera agregar alguna otra sección adicional está bien.
- Presentación oral: Realizar una presentación: la estructura mínima debe ser la siguiente: Tabla de contenido, Introducción, Metodología, Descripción del corpus, Resultados, Conclusiones. No se olviden de incluir el número de páginas en las diapositivas.
- Tanto la presentación como el reporte se debe entregar en formato PDF en la plataforma virtual.
- Código documentado y reproducible

**Se requiere entregar proyecto para tener derecho al primer final. Si no se entrega el proyecto, se presentará sólo el segundo final.**