



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN LINGÜÍSTICA



Propuesta de programa para impartir asignaturas en

(INDICAR EL PROGRAMA QUE CORRESPONDA)

NB No se considerarán propuestas que no incluyan toda la información solicitada

MAESTRÍA EN LINGÜÍSTICA APLICADA

MAESTRÍA EN LINGÜÍSTICA HISPÁNICA

TÍTULO DE LA ASIGNATURA

Aproximaciones contemporáneas en la lingüística computacional

PROFESOR/A

Dra. María Ximena Gutiérrez Vasques

CLAVE	SEMESTRE	CAMPO DE CONOCIMIENTO		NÚMERO DE CRÉDITOS
o Optativa		HORAS		HORAS AL SEMESTRE
		HORAS POR SEMANA		4
		36		64
o Curso		MODALIDAD	TEORÍA	PRÁCTICA
			36	28
		DURACIÓN DEL CURSO		semestral

ACTIVIDAD ACADÉMICA CON SERIACIÓN SUBSECUENTE

ACTIVIDAD ACADÉMICA CON SERIACIÓN ANTECEDENTE

OBJETIVO(S) DEL CURSO O SEMINARIO*Objetivo general*

Introducir a los estudiantes en los métodos actuales de la lingüística computacional y el procesamiento del lenguaje natural (PLN), con énfasis en modelos de lenguaje probabilísticos y otros métodos actuales. Se busca explorar cómo estas técnicas permiten procesar, analizar y generar lenguaje natural, y qué tipos de hipótesis lingüísticas podrían investigarse a partir de su funcionamiento.

Objetivos específicos

- Contextualizar el desarrollo reciente de la lingüística computacional y su impacto en el estudio del lenguaje
- Explorar enfoques situados en la intersección entre la lingüística, la estadística y las ciencias de la computación
- Desarrollar habilidades para analizar bases de datos lingüísticas utilizando herramientas computacionales
- Fomentar una mirada crítica, desde la lingüística, sobre las limitaciones y posibilidades de las tecnologías actuales de generación automática de texto

ÍNDICE TEMÁTICO

			Horas
UNIDAD	TEMA	Teóricas	Prácticas
1	Introducción a la lingüística computacional: De lo formal a lo estadístico	12	4
2	Los grandes modelos del lenguaje (LLMs): Fundamentos y funcionamiento	8	8
3	Enfoques cuantitativos en tipología y diversidad lingüística: Aplicaciones con herramientas de lingüística computacional	4	12
4	Discusiones actuales: Análisis crítico de investigaciones en la intersección entre lingüística y modelos de lenguaje	12	4
TOTAL DE HORAS			36
SUMA TOTAL DE HORAS:			64

CONTENIDO TEMÁTICO DESGLOSADO

UNIDAD	TEMA/SUBTEMAS
1.	Introducción a la lingüística computacional: De lo formal a lo estadístico 1.1 El lenguaje natural como desafío central de la inteligencia artificial 1.2 Desarrollo histórico. Orígenes de la lingüística computacional 1.3 Enfoques formales en el análisis del lenguaje 1.4 Métodos estadísticos y el uso de la web como corpus 1.5 Aproximaciones computacionales a los niveles lingüísticos: fonología, morfología, sintaxis, semántica y pragmática 1.6 Métodos actuales de tokenización de palabras ortográficas
2.	Los grandes modelos del lenguaje (LLMs): Fundamentos y funcionamiento 2.1 ¿Qué es un modelo del lenguaje estadístico/probabilístico? 2.2 Semántica distribucional como base de los modelos del lenguaje actuales 2.3 Características de los principales asistentes generativos de texto comerciales 2.4 Familiarización con Python y programación asistida por IA
3.	Enfoques cuantitativos en tipología y diversidad lingüística: Aplicaciones con herramientas de lingüística computacional 3.1 Exploración automática de bases de datos tipológicas y genealógicas: WALS, Grambank, Glottolog 3.2 Panorama de la diversidad lingüística en las tecnologías del lenguaje actuales 3.4 Diversidad y sesgos en IA 3.4 Otras aplicaciones
4.	Discusiones actuales: Análisis crítico de investigaciones en la intersección entre lingüística y modelos de lenguaje 4.1 Lecturas del estado del arte en distintas temáticas: <ul style="list-style-type: none"> - Características lingüísticas codificadas en los modelos de redes neuronales artificiales ¿qué aprenden realmente estos modelos? - Modelado del diálogo en los LLMs - Aplicaciones de LLMs al análisis del discurso 4.2 Limitaciones actuales en los modelos de lenguaje 4.3 Discusión final y prospectivas

BIBLIOGRAFÍA BÁSICA

MARCO BARONI. (2022). ON THE PROPER ROLE OF LINGUISTICALLY ORIENTED DEEP NET ANALYSIS IN LINGUISTIC THEORISING. IN ALGEBRAIC STRUCTURES IN NATURAL LANGUAGE. CRC PRESS

JURAFSKY, D., & MARTIN, J. H. (2025). SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS, AND SPEECH RECOGNITION WITH LANGUAGE MODELS (3RD ED.) [ONLINE MANUSCRIPT]. [HTTPS://WEB.STANFORD.EDU/~JURAFSKY/SLP3/](https://web.stanford.edu/~jurafsky/SLP3/)

MITKOV, R. (ED.). (2022). THE OXFORD HANDBOOK OF COMPUTATIONAL LINGUISTICS. OXFORD UNIVERSITY PRESS.

BIBLIOGRAFÍA COMPLEMENTARIA

Gutierrez-Vasques, X., Bentz, C., & Samardžić, T. (2023). Languages through the looking glass of BPE compression. *Computational Linguistics*, 49(4), 943-1001.

Miaschi, A., Dell'Orletta, F., & Venturi, G. (2024, November). Evaluating large language models via linguistic profiling. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 2835-2848).

Hayati, S., Lee, M., Rajagopal, D., & Kang, D. (2024, November). How Far Can We Extract Diverse Perspectives from Large Language Models?. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 5336-5366).

Guo, R., Nejadgholi, I., Dawkins, H., Fraser, K. C., & Kiritchenko, S. (2024, November). Adaptable Moral Stances of Large Language Models on Sexist Content: Implications for Society and Gender Discourse. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 19548-19564).

Tian, Y., Huang, T., Liu, M., Jiang, D., Spangher, A., Chen, M., ... & Peng, N. (2024). Are Large Language Models Capable of Generating Human-Level Narratives?. *arXiv preprint arXiv:2407.13248*.

Ou, J., Lu, J., Liu, C., Tang, Y., Zhang, F., Zhang, D., & Gai, K. (2024, June). DialogBench: Evaluating LLMs as Human-like Dialogue Systems. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Volume 1: Long

Papers) (pp. 6137-6170).

Bentz, C., Gutierrez-Vasques, X., Sozinova, O., & Samardžić, T. (2023). Complexity trade-offs and equi-complexity in natural languages: a meta-analysis. *Linguistics Vanguard*, 9(s1), 9-25.

ESTRATEGIAS DIDÁCTICAS		MECANISMOS DE EVALUACIÓN DEL APRENDIZAJE DE LOS ALUMNOS			
Exposición oral por parte del profesor	()	Exámenes Parciales	()		
Exposición audiovisual a cargo del profesor	(X)	Examen final escrito	()		
Exposición oral por parte de los alumnos	()	Trabajos y tareas fuera del aula	(X)		
Ejercicios dentro de clase	(X)	Exposición en seminarios por los alumnos	(X)		
Ejercicios fuera del aula	()	Participación en clase	(X)		
Lecturas obligatorias	(X)	Asistencia	()		
Trabajo de Investigación	()	Seminario	()		
Otro		Trabajo de investigación	(X)		
NB: En los cursos obligatorios o que constituyen el primer acercamiento a un tema, la presentación de los temas debe descansar preponderantemente en el profesor.		Otro			
CRITERIOS DE EVALUACIÓN					
La evaluación será continua e incluirá:					
a) Ejercicios prácticos y dinámicas grupales en clase					
b) Participación activa en clase					
c) Exposición por parte de los alumnos en la Unidad 4					
d) Trabajo de investigación breve sobre un tema relacionado con el curso					
PERFIL PROFESIOGRÁFICO					
El docente deberá contar con una formación interdisciplinaria en lingüística computacional y tecnologías del lenguaje. Deberá manejar con solvencia las técnicas y conceptos propios de las tecnociencias, así como comprender la relevancia de estos enfoques para el estudio del lenguaje natural.					

PROPUESTAS DE HORARIO PARA IMPARTIR EL CURSO:

1.- Martes en cualquier horario

2- Lunes en cualquier horario

FECHA: **22 DE MAYO DEL 2025**