

**FICHA IDENTIFICATIVA****DATOS DE LA ASIGNATURA****Código:** 46757**Nombre:** Metodología para la investigación lingüística hispánica II: lingüística de corpus y metodologías cua**Ciclo:** Máster Universitario Oficial**Créditos ECTS:** 6**Curso académico:** 2025-26**TITULACIONES**

Titulación	Centro	Curso	Periodo
2267 - Máster Universitario en Estudios Hispánicos Avanzados	Facultat de Filologia, Traducció i Comunicació	1	Anual

MATERIAS

Titulación	Materia	Carácter
2267 - Máster Universitario en Estudios Hispánicos Avanzados	Metodología para la investigación lingüística hispánica II: lingüística de corpus y metodologías cua	OPTATIVA

COORDINACIÓN

ALBELDA MARCO MARTA

CABEDO NEBOT ADRIAN

RESUMEN

En el plan de estudios del Máster, la asignatura se integra en el Itinerario I (Lengua Española) y en el Itinerario III (Formación Hispanística), orientación investigadora.

Los contenidos de la materia se organizan en dos bloques temáticos (impartidos en horario distinto):

En el bloque temático *Lingüística de corpus* se aborda la elaboración y el tratamiento de corpus orales y escritos, necesarios en la realización de una investigación en el área de Lengua Española. Aunque su perfil tiene una base teórica al describir el concepto y tipología de corpus y la disciplina correspondiente (Lingüística del corpus), se completa con un componente práctico esencial dirigido hacia el diseño de un corpus y tratamiento informático de los datos obtenidos. Además, se revisan los corpus actuales en español para su empleo en la investigación.

En el bloque temático (*usos de técnicas estadísticas*) la asignatura se centra en la aplicación práctica de



herramientas fundamentales para el tratamiento cuantitativo de datos en la investigación lingüística hispánica. Se dirigirá a los alumnos la utilización de programas informáticos de especial interés en la investigación lingüística, fundamentalmente orientados a la práctica estadística

Tal y como se detalla en el volumen de trabajo, las clases teóricas se complementan con clases prácticas. A estas actividades formativas de carácter presencial se añade la participación en conferencias y seminarios organizados por el Departamento de Filología Española.

La docencia del bloque *Lingüística de corpus* (temas 1, 2 y 3) será impartida, generalmente, en el primer cuatrimestre.

La docencia del bloque *Usos de técnicas estadísticas* (temas 4, 5 y 6) se desarrollará, en principio, en el segundo cuatrimestre.

En esta asignatura se atiende especialmente a los siguientes ODS: ODS4, ODS8 y ODS9.

CONOCIMIENTOS PREVIOS

RELACIÓN CON OTRAS ASIGNATURAS DE LA MISMA TITULACIÓN

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

OTROS TIPOS DE REQUISITOS

COMPETENCIAS / RESULTADOS DE APRENDIZAJE

-

Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

Que los/las estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

Que los/las estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

Que los/las estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Ser capaces de elaborar un proyecto en el ámbito de estudio del máster y presentar sus resultados, mediante un discurso elaborado y coherente.



Ser capaces de emplear los recursos institucionales, los grandes instrumentos científicos, las técnicas, las bases de datos y bibliotecas, la bibliografía, las revistas especializadas y los nuevos recursos proporcionados por las TIC en el ámbito de estudio.

DESCRIPCIÓN DE CONTENIDOS

1. Lingüística de corpus y lingüística con corpus

Qué es y por qué la lingüística de corpus. Qué es un corpus lingüístico. Corpus-based approach vs. The corpus-driven approach. Características de los corpus y tipos de corpus. Otras técnicas de elicitación de datos. Aplicaciones a la investigación lingüística: análisis del acceso y empleo de corpus en artículos científicos.

2. Panorama de los corpus hispánicos

Corpus de concordancias y corpus de acceso al texto o discurso; escritos y orales, con fines generales y con fines específicos. Corpus para tecnologías del habla, corpus paralelos, corpus de lenguajes técnicos, corpus orientados a ASL.

3. La confección de un corpus: DIY

Selección de la muestra. Representatividad, tamaño, adecuación a los fines. Técnicas de transcripción. Programas para la transcripción. Anotado, etiquetado, lematizado. Copyright.

4. Bases del análisis estadístico

Tipología de variables. Variables ocultas. Población y muestra. Niveles de significación: valor p.



5. Introducción básica al manejo del programa R

Data frames. Variables. Ventanas de navegación.

6. Pruebas estadísticas de visualización y contraste poblacional

Técnicas de visualización: gráficos de barras, diagramas de caja, motion charts. Mosaicplot, chi cuadrado y residuos estandarizados. Student T-Test / Pruebas ANOVA. Análisis múltiple de correspondencias. Árbol de decisiones.

Estos contenidos se plasmarán en los siguientes resultados de aprendizaje:

Ser capaces de identificar y delimitar los objetivos de una investigación en el ámbito de estudio de la lingüística hispánica, para evaluar su relevancia científica, para formular las preguntas guía y detectar las cuestiones poco o mal estudiadas y para dotarse de los medios para llevarla a cabo.

Ser capaces de elaborar y catalogar corpus, eligiendo la metodología específica en función del público potencial al que se dirigen.

Ser capaces de emplear los recursos institucionales y los grandes instrumentos científicos, las bases de datos y otros recursos tecnológicos en el ámbito del estudio y análisis de corpus en el marco de la investigación lingüística hispánica.

Ser capaces de emplear los nuevos recursos proporcionados por las TIC, en el ámbito de estudio de la investigación lingüística hispánica.

VOLUMEN DE TRABAJO (HORAS)

ACTIVIDADES PRESENCIALES

Actividad	Horas
Teoría-Prácticas	40,00
Total horas	40,00

**ACTIVIDADES NO PRESENCIALES**

Actividad	Horas
Asistencia a otras actividades	6,00
Elaboración de trabajos individuales o en grupo	0,00
Estudio y trabajo autónomo	104,00
Preparación de clases	0,00
Preparación de actividades de evaluación	0,00
Resolución de casos prácticos	0,00
Total horas	110,00

METODOLOGÍA DOCENTE

Lección magistral (exposición de contenidos a partir de desarrollos teóricos) con diferentes recursos (textos, audiovisuales, etc.).

Clases prácticas participativas (ejercicios, exposiciones, debates, test, etc.), realizadas en grupo y/o de forma individual.

Usos de procedimientos y estrategias deductivos en la aplicación del conocimiento adquirido: detección y resolución de problemas, actividades de forma individual o en grupo (ejercicios, lectura y comentario crítico de diferentes textos, realización de trabajos escritos y/o exposición oral de los resultados, etc.) a partir de la exposición previa de contenidos.

Usos de procedimientos y estrategias inductivos en la aplicación del conocimiento adquirido: detección y resolución de problemas, actividades de forma individual o en grupo (ejercicios, lectura y comentario crítico de diferentes textos, realización de trabajos escritos y/o exposición oral de los resultados, etc.) a partir de directrices básicas para formular un desarrollo teórico posterior.

sarrollo teórico posterior.p>

EVALUACIÓN

a) Valoración de la participación activa en el aula (exposición y ejercicio): hasta 2 puntos

- Presentación oral en clase sobre análisis crítico de metodologías basadas en corpus y ejercicio de investigación de corpus. Hasta 1 punto.
- Cuestionario de estadística básica y de uso de R para el análisis de datos lingüísticos. Hasta 1 punto.

NO recuperables en segunda convocatoria (actividad presencial en el aula). Se mantendrá la nota en la segunda convocatoria.



b) Valoración del trabajo escrito, ejercicio y cuestionario, en horas no presenciales. Hasta 8 puntos:

- Trabajo escrito. Hasta 4 puntos.
- Trabajo escrito. Hasta 4 puntos.

Recuperables en segunda convocatoria.

OBSERVACIÓN IMPORTANTE: Para poder aprobar la asignatura y para que se puedan sumar las notas de las dos partes (a y b) del sistema de evaluación, será necesario que en las tareas de la parte no presencial (trabajo, cuestionario y ejercicio) se obtengan, en su conjunto, al menos 5 puntos.

Tanto en las pruebas como en las actividades se exigirá corrección ortográfica, gramatical y léxica. Cada incorrección comportará una reducción de la calificación obtenida, que puede llegar hasta el suspenso.

La honestidad intelectual es vital en las comunidades académicas, y para la justa evaluación del trabajo del estudiantado. Todos los trabajos presentados en este curso han de ser de autoría original. No se admitirán trabajos en los que se haga uso de colaboración fraudulenta o la composición con la ayuda de inteligencia artificial (ChatGPT u otros), excepto si su utilización forma parte de los contenidos de la asignatura y está autorizada por el profesorado que la imparte.

La evaluación y el sistema general de calificaciones seguirán la normativa de la Universitat de València aprobada por el Consejo de Gobierno el 30/5/2017-ACGUV108/2017.

BIBLIOGRAFÍA

Básica

BRIZ, Antonio y Grupo VAL.ES.CO (2002): *Corpus de conversaciones coloquiales*, Madrid, Arco.

BRIZ, Antonio y Grupo VAL.ES.CO (2002): «La transcripción de la lengua hablada: el sistema del Grupo Val.Es.Co», *Español Actual*, 77-78, 57-85.

BRIZ, Antonio y Marta ALBELDA (2009): «Estado actual de los corpus de lengua española hablada y escrita: I+D». *Anuario del Instituto Cervantes*, Instituto Cervantes, 165-226 [en línea]: http://cvc.cervantes.es/lengua/anuario/anuario_09/briz_albeida/p01.htm.

ENGHELS, Renata; Clara VANDERSCHUEREN y Miriam BOUZOUITA (2015): «Panorama de los corpus y textos del español peninsular contemporáneo». En *Manual of Romance Linguistics*, 7, pp. 147-170.



GRIES, S. Th. (2021). *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter. <https://doi.org/10.1515/9783110718256>

LEVSHINA, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.

MCENERY, Tonny, Richard XIAO y Yukio TONO (2006): *Corpus Based Language Studies*, Routledge, London [en línea]: <http://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xCBLS/CBLS.htm>

O'KEEFFE, A. y M. MCCARTHY (2010): *The Routledge Handbook of Corpus Linguistics*. Routledge.

TORRUELLA, Joan y Joaquim LLISTERRI (1999): «Diseño de corpus textuales y orales», en Bleuca, J. M. et alii. (eds.): *Filología e Informàtica. Nuevas tecnologías en los estudios filológicos*, Universidad Aut3noma de Barcelona, pp. 45-77 [en línea]: http://liceu.uab.es/%7Ejoaquim/publicacions/Torruella_Llisterri_99.pdf.

Complementaria

AIJMER, Karin, ed. (2009): *Corpora and Language Teaching*, Amsterdam, John Benjamins.

CARAVEDO, Rocío (1999): *Lingüística del corpus. Cuestiones teórico-metodológicas aplicadas al español*, Universidad de Salamanca.

CRUZ PIÑOL, Mar (2012): *Lingüística de corpus y adquisición del español como 2/L*, Madrid, Arco.

PAYRATÓ, Lluís et alii (1996): *Corpus, corpora*, Barcelona, PPU.

PÉREZ, Chantal (2002): «Explotación de los c3rpora textuales informatizados para la creaci3n de bases de datos terminol3gicas basadas en el conocimiento», *Estudios de Lingüística del Espa3ol (ELiEs)*, vol.18 [en línea]: <http://elies.rediris.es/elies18/index.html>.

Se indicarán más referencias al inicio de curso. / S'indicaran més referències a l'inici de curs. / More references will be indicated at the beginning of the course.