

**FITXA IDENTIFICATIVA****DADES DE L'ASSIGNATURA****Codi:** 46757**Nom:** Metodologia per a la investigació lingüística hispànica II: lingüística de corpus i metodologies qua**Cicle:** Màster Universitari Oficial**Crèdits ECTS:** 6**Curs acadèmic:** 2026-27**TITULACIONS**

Titulació	Centre	Curs	Període
2267 - M.U. Estudis Hispànics Avançats: Aplicacions i Investigació	Facultat de Filologia, Traducció i Comunicació	1	Anual

MATÈRIES

Titulació	Matèria	Caràcter
2267 - M.U. Estudis Hispànics Avançats: Aplicacions i Investigació	Metodologia per a la investigació lingüística hispànica II: lingüística de corpus i metodologies qua	OPTATIVA

COORDINACIÓ

ALBELDA MARCO MARTA

CABEDO NEBOT ADRIAN

RESUM

En el pla d'estudis del Màster, l'assignatura s'integra en l'itinerari I (Llengua Espanyola) i en l'itinerari III (Formació Hispanística), orientació investigadora.

Els continguts de la matèria s'organitzen en dos blocs temàtics (impartits en horari diferent):

En el bloc temàtic Lingüística de corpus s'aborda l'elaboració i el tractament de corpus orals i escrits, necessaris per a la realització d'una investigació en l'àrea de Llengua Espanyola. Tot i que el seu perfil té una base teòrica en descriure el concepte i la tipologia de corpus i la disciplina corresponent (Lingüística del corpus), es completa amb un component pràctic essencial dirigit cap al disseny d'un corpus i el tractament informàtic de les dades obtingudes. A més, es revisen els corpus actuals en espanyol per al seu ús en la investigació.

En el bloc temàtic Usos de tècniques estadístiques l'assignatura se centra en l'aplicació pràctica d'eines fonamentals per al tractament quantitatiu de dades en la investigació lingüística hispànica. Es dirigirà als



46757 Metodologia per a la investigació lingüística hispànica II: lingüística de corpus i metodologies qua

alumnes en l'ús de programes informàtics d'especial interès en la investigació lingüística, principalment orientats a la pràctica estadística.

Segons es detalla en el volum de treball, les classes teòriques es complementen amb classes pràctiques. A aquestes activitats formatives de caràcter presencial s'afegeix la participació en conferències i seminaris organitzats pel Departament de Filologia Espanyola.

La docència del bloc Lingüística de corpus (temes 1, 2 i 3) s'impartirà, generalment, en el primer quadrimestre.

La docència del bloc Usos de tècniques estadístiques (temes 4, 5 i 6) es desenvoluparà, en principi, en el segon quadrimestre.

En aquesta assignatura s'atén especialment els següents ODS: ODS4, ODS8 i ODS9.

CONEXIMENTS PREVIS

RELACIÓ AMB ALTRES ASSIGNATURES DE LA MATEIXA TITULACIÓ

No s'ha especificat restriccions de matrícula amb altres assignatures del pla d'estudis.

ALTRES TIPUS DE REQUISITS

COMPETÈNCIES / RESULTATS D' APRENTATGE

2267 - M.U. Estudis Hispànics Avançats: Aplicacions i Investigació

Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i / o aplicació d'idees, sovint en un context de recerca.

Que els estudiants posseïsquen les habilitats d'aprenentatge que els permeten continuar estudiant d'una forma que haurà de ser en gran manera autodirigida o autònoma.

Que els estudiants sàpiguen comunicar les conclusions (i els coneixements i les raons últimes que les sustenten) a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Que els estudiants siguin capaços d'integrar coneixements i afrontar la complexitat de formular judicis a partir d'una informació que, sent incompleta o limitada, incloga reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

Ser capaços d'elaborar un projecte en l'àmbit d'estudi del màster i presentar els seus resultats, per mitjà d'un discurs elaborat i coherent.



Ser capaços d'emprar els recursos institucionals, els grans instruments científics, les tècniques, les bases de dades i biblioteques, la bibliografia, les revistes especialitzades i els nous recursos proporcionats per les TIC en l'àmbit d'estudi.

DESCRIPCIÓ DE CONTINGUTS

1. Lingüística de corpus i lingüística amb corpus

Què és i per què la lingüística de corpus. Què és un corpus lingüístic. Aproximació basada en corpus vs. Aproximació impulsada pel corpus. Característiques dels corpus i tipus de corpus. Altres tècniques d'elicitació de dades. Aplicacions en la investigació lingüística: anàlisi de l'accés i ús de corpus en articles científics.

2. Panorama dels corpus hispànics

Corpus de concordàncies i corpus d'accés al text o discurs; escrits i orals, amb finalitats generals i amb finalitats específiques. Corpus per a tecnologies del parlament, corpus paral·lels, corpus de llenguatges tècnics, corpus orientats a ASL.

3. La confecció d'un corpus: DIY

Selecció de la mostra. Representativitat, grandària, adequació als objectius. Tècniques de transcripció. Programari per a la transcripció. Anotat, etiquetat, lematitzat. Drets d'autor.

4. Bases de l'anàlisi estadístic

Tipologia de variables. Variables ocultes. Població i mostra. Nivells de significació: valor p.



5. Introducció bàsica al maneig del programa R

Data frames. Variables. Finestres de navegació.

6. Proves estadístiques de visualització i contrast poblacional

Tècniques de visualització: gràfics de barres, diagrames de caixa, motion charts. Mosaicplot, chi quadrat i residus estandarditzats. Student T-Test / Proves ANOVA. Anàlisi múltiple de correspondències. Arbre de decisions.

Aquests continguts es reflectiran en els següents resultats d'aprenentatge:

Ser capaços d'identificar i delimitar els objectius d'una investigació en l'àmbit d'estudi de la lingüística hispanica, per avaluar la seua rellevància científica, per formular les preguntes guia i detectar les qüestions poc o mal estudiades, i per dotar-se dels mitjans per dur-la a terme.

Ser capaços d'elaborar i catalogar corpus, triant la metodologia específica en funció del públic potencial al qual es dirigeixen.

Ser capaços d'emplear els recursos institucionals i els grans instruments científics, les bases de dades i altres recursos tecnològics en l'àmbit de l'estudi i anàlisi de corpus dins del marc de la investigació lingüística hispanica.

Ser capaços d'emplear els nous recursos proporcionats per les TIC en l'àmbit de l'estudi de la investigació lingüística hispanica.

VOLUM DE TREBALL (HORES)

**ACTIVITATS PRESENCIALS**

Activitat	Hores
Teoria-Pràctiques	40,00
Total hores	40,00

ACTIVITATS NO PRESENCIALS

Activitat	Hores
Assistència a altres activitats	6,00
Elaboració de treballs individuals o en grup	0,00
Estudi i treball autònom	104,00
Preparació de classes	0,00
Preparació d'activitats d'avaluació	0,00
Resolució de casos pràctics	0,00
Total hores	110,00

METODOLOGIA DOCENT

Lliçó magistral (exposició de continguts a partir de desenvolupaments teòrics) amb diferents recursos (textos, audiovisuals, etc.).

Classes pràctiques participatives (exercicis, exposicions, debats, proves, etc.), realitzades en grup i/o de forma individual.

Ús de procediments i estratègies deductius en l'aplicació del coneixement adquirit: detecció i resolució de problemes, activitats de manera individual o en grup (exercicis, lectura i comentari crític de diferents textos, realització de treballs escrits i/o exposició oral dels resultats, etc.) a partir de l'exposició prèvia de continguts.

Ús de procediments i estratègies inductius en l'aplicació del coneixement adquirit: detecció i resolució de problemes, activitats de manera individual o en grup (exercicis, lectura i comentari crític de diferents textos, realització de treballs escrits i/o exposició oral dels resultats, etc.) a partir de directrius bàsiques per a formular un desenvolupament teòric posterior.

AVALUACIÓ

a) Valoració de la participació activa en l'aula (exposició i exercici): fins a 2 punts:

Presentació oral a classe sobre anàlisi crític de metodologies basades en corpus i exercici d'investigació de corpus. Fins a 1 punt.

Qüestionari d'estadística bàsica i ús de R per a l'anàlisi de dades lingüístiques. Fins a 1 punt.

No recuperables en segona convocatòria (activitat presencial a l'aula). Es mantindrà la nota en la segona



convocatòria.

b) Valoració del treball escrit, exercici i qüestionari, en hores no presencials. Fins a 8 punts:

Treball escrit. Fins a 4 punts.

Treball escrit. Fins a 4 punts.

Recuperables en segona convocatòria.

OBSERVACIÓ IMPORTANT: Per a poder aprovar l'assignatura i perquè es puguin sumar les notes de les dues parts (a i b) del sistema d'avaluació, serà necessari que en les tasques de la part no presencial (treball, qüestionari i exercici) s'obtinguen, en conjunt, almenys 5 punt.

Tant en les proves com en les activitats, s'exigirà correcció ortogràfica, gramatical i lèxica. Cada incorrecció comportarà una reducció de la qualificació obtinguda, que pot arribar a suposar un suspens.

L'avaluació i el sistema general de qualificacions seguiran la normativa de la Universitat de València aprovada pel Consell de Govern el 30/5/2017-ACGUV108/2017.

L'honestedat intel·lectual és vital en les comunitats acadèmiques, i per a la justa avaluació del treball de l'estudiantat. Tots els treballs presentats en aquest curs han de ser d'autoria original. No s'admetran treballs en els quals es faci ús de col·laboració fraudulenta o la composició amb l'ajuda d'intel·ligència artificial (ChatGPT o altres), excepte si la seua utilització forma part dels continguts de l'assignatura i està autoritzada pel professorat que la imparteix.

BIBLIOGRAFIA

Bàsica

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 línia]: 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 Blecua, J. M. et alii. (eds.): *Filología e Informática. Nuevas tecnologías en los estudios filológicos*, Universidad Autónoma de Barcelona, pp. 45-77 [en línea]: http://liceu.uab.es/%7Ejoaquim/publicacions/Torruella_Llisterri_99.pdf.

Complementària

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 corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento», *Estudios de Lingüística del Español (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.



VNIVERSITAT DE VALÈNCIA

Guia Docent
46757 Metodologia per a la investigació lingüística
hispanica II: lingüística de corpus i metodologies qua
