



COURSE DATA

DATA SUBJECT

Code: 46757
Name: Metodología para la investigación lingüística hispánica II: lingüística de corpus y metodologías cua
Cycle: Master's Degree
ECTS Credits: 6
Academic year: 2025-26

STUDY (S)

Degree	Center	Acad. year	Period
2267 - Master's Degree in Advanced Hispanic Studies: Application and Research	Facultat de Filologia, Traducció i Comunicació	1	Annual

SUBJECT-MATTER

Degree	Subject-matter	Character
2267 - Master's Degree in Advanced Hispanic Studies: Application and Research	Metodología para la investigación lingüística hispánica II: lingüística de corpus y metodologías cua	ELECTIVES

COORDINATION

ALBELDA MARCO MARTA

CABEDO NEBOT ADRIAN

SUMMARY

In the curriculum of the Master's Degree, the subject is integrated into Itinerary I (Spanish Language) and Itinerary III (Hispanic Training), research orientation.

The contents of the subject are organized into two thematic blocks (taught at different times):

The thematic block *Corpus Linguistics* addresses the elaboration and treatment of oral and written corpora, necessary in the conduct of research in the area of Spanish Language. Although its profile has a theoretical basis when describing the concept and typology of corpus and the corresponding discipline (Linguistics of the corpus), it is completed with an essential practical component aimed at the design of a corpus and computer processing of the data obtained. In addition, the current corpora in Spanish are reviewed for use in the research.

In the thematic block (*uses of statistical techniques*) the subject focuses on the practical application of fundamental tools for the quantitative processing of data in Hispanic linguistic research. Students will be



directed to use computer programs of special interest in linguistic research, fundamentally oriented to statistics.

As detailed in the workload, theoretical classes are complemented by practical classes. These face-to-face training activities are complemented by attendance to conferences and seminars organized by the Department of Spanish Philology.

The teaching of the *Corpus Linguistics* block (topics 1, 2 and 3) will generally be given in the first semester.

The teaching of the *Uses of Statistical Techniques* block (topics 4, 5 and 6) will be developed, in principle, in the second semester.

In this subject the following SDGs are especially addressed: SDG4, SDG8 i SDG9.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

There are no specified enrollment restrictions with other subjects of the curriculum.

OTHER REQUIREMENTS

COMPETENCES / LEARNING OUTCOMES

-

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.

Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.

Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.

Students should demonstrate self-directed learning skills for continued academic growth.

Students should possess and understand foundational knowledge that enables original thinking and



research in the field.

DESCRIPTION OF CONTENTS

1. Corpus Linguistics and Corpus Linguistics

What is corpus linguistics and why is it used. What is a linguistic corpus? A corpus-based approach. The corpus-driven approach. Characteristics of corpora and types of corpora. Other data collection techniques Applications to linguistic research: analysis of access and use of corpora in scientific articles.

2. Overview of Hispanic corpora

Concordance corpora and corpus of access to the text or speech; written and oral, for general purposes and for specific purposes. Corpus for speech technologies, parallel corpora, corpus of technical languages, ASL oriented corpora.

3. The making of a corpus: DIY

Sample selection. Representativeness, size, suitability for the purposes. Transcription techniques. Programs for transcription. Annotating, labelling, lemmatisation. Copyright.

4. The bases of statistical analysis

Typology of variables. Hidden variables. Population and sample. Levels of significance: p-value

5. Basic Introduction to R Project Management

Data frames. Frames Variables. Navigation windows.



6. Statistical tests of visualization and population contrast

Visualization techniques: bar charts, boxplots, motion charts. Mosaicplot, chi-square and standardized waste. Student T-Test /ANOVA Tests. Correspondence analysis Decision tree.

These contents will be reflected in the following learning outcomes:

To be able to identify and define the objectives of a research in the field of study of Hispanic linguistics, to evaluate its scientific relevance, to formulate the guiding questions and detect the little or poorly studied issues and to equip themselves with the means to carry it out.

To be able to elaborate and catalogue corpus, choosing the specific methodology depending on the potential audience to which they are addressed.

To be able to use institutional resources and major scientific tools, databases and other technological resources in the field of corpus study and analysis within the framework of Hispanic linguistic research.

To be able to use the new resources provided by ICT, in the field of study of Hispanic linguistic research.

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theoretical and practical classes	40,00
Total hours	40,00

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	6,00
Individual or group project	0,00
Independent study and work	104,00
Preparation of lessons	0,00
Preparation for assessment activities	0,00



Resolution of case studies	0,00
Total hours	110,00

TEACHING METHODOLOGY

Lectures (presentation of content based on theoretical developments) using various resources (texts, audiovisual materials, etc.).

Participatory practical classes (exercises, presentations, debates, tests, etc.), conducted individually or in groups.

Use of deductive procedures and strategies in applying acquired knowledge: problem detection and resolution, activities conducted individually or in groups (exercises, reading and critical commentary on various texts, writing assignments and/or oral presentation of results, etc.) based on prior content presentation.

Use of inductive procedures and strategies in applying acquired knowledge: problem detection and resolution, activities conducted individually or in groups (exercises, reading and critical commentary on various texts, writing assignments and/or oral presentation of results, etc.) based on basic guidelines to formulate subsequent theoretical development.

EVALUATION

a) Assessment of active participation in the classroom (presentation and exercise): up to 2 points:

- Oral presentation in class on critical analysis of corpus-based methodologies and corpus research exercise. Up to 1 point.
- Questionnaire on basic statistics and the use of R for the analysis of linguistic data. Up to 1 point.

Not recoverable on second call (face-to-face activity in the classroom). The grade will be kept in the second call.

b) Assessment of the written work, exercise and questionnaire, in non-face-to-face hours. Up to 8 points:



- Written assignment Up to 4 points.
- Written assignment Up to 4 points.

Recoverable on second call.

IMPORTANT OBSERVATION: In order to pass the subject and so that the grades of the two parts a) and b) of the evaluation system can be added, it will be necessary that in the tasks of the non-face-to-face part (work, questionnaire and exercise) at least 5 points are obtained, as a whole.

Both in examinations and coursework, orthographic, grammatical, and lexical accuracy will be required. Each spelling or language error will result in a reduction of the final mark, which may lead to a failing grade.

Intellectual honesty is vital in academic communities, and for the fair evaluation of student work. All work submitted in this course must be original authorship. Papers that make use of fraudulent collaboration or composition with the help of artificial intelligence (ChatGPT or others) will not be accepted, except if their use is part of the course content and is authorized by the faculty teaching the course.

The evaluation and the general grading system will follow the regulations of the University of Valencia approved by the Governing Council on 30/5/2017-ACGUV108/2017.

REFERENCES

Basic

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 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.

Additional

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.