

Tratamiento y Análisis de Datos Lingüísticos

Código: 45505

Créditos ECTS: 5

2025/2026

Titulación	Tipo	Curso
Advanced Studies in Catalan Language and Literature	OP	1

Contacto

Nombre: Gemma Repiso Puigdelliura

Correo electrónico: gemma.repiso@uab.cat

Equipo docente

(Externo) Josefina Carrera-Sabaté

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

No hay prerequisitos.

Objetivos y contextualización

La asignatura tiene como objetivo principal proporcionar al estudiante las herramientas metodológicas necesarias para llevar a cabo una investigación rigurosa en el ámbito de la lingüística, con un énfasis especial en el estudio de la variación lingüística centrada en el componente fónico de la lengua. Se proporcionarán conocimientos sobre el ciclo completo de una investigación lingüística enmarcada en la sociolingüística de la variación: desde la definición del objeto de estudio y la recogida de datos, pasando por la observación cualitativa y cuantitativa hasta el análisis estadístico y la formalización de los resultados. Se impartirán conocimientos sobre el software R, que se utilizará como herramienta principal para el tratamiento, la organización y la visualización de datos lingüísticos, así como para la aplicación de técnicas estadísticas descriptivas e inferenciales que permitan extraer conclusiones a partir de datos empíricos.

El estudiante deberá ser capaz de:

- Identificar y describir fenómenos de variación lingüística.
- Aplicar métodos de investigación cuantitativos y cualitativos en el análisis de datos lingüísticos.
- Utilizar herramientas de tratamiento de datos, especialmente el software R, para la manipulación y visualización de datos lingüísticos.
- Formular hipótesis, realizar contrastes estadísticos, analizar resultados y extraer conclusiones significativas a partir de datos empíricos.
- Presentar una propuesta de investigación y un informe de análisis de datos con estructura científica y claridad expositiva.

Resultados de aprendizaje

1. CA23 (Competencia) Aplicar los conocimientos, métodos y herramientas de recogida y tratamiento de datos para la creación de corpus lingüísticos.
2. CA23 (Competencia) Aplicar los conocimientos, métodos y herramientas de recogida y tratamiento de datos para la creación de corpus lingüísticos.
3. CA24 (Competencia) Examinar el diseño de una investigación lingüística, los procesos de selección de informantes y las técnicas de recogida de datos lingüísticos siguiendo los principios de la ética investigadora.
4. CA24 (Competencia) Examinar el diseño de una investigación lingüística, los procesos de selección de informantes y las técnicas de recogida de datos lingüísticos siguiendo los principios de la ética investigadora.
5. KA30 (Conocimiento) Identificar las fases de la recogida, tratamiento, análisis, formalización y presentación de datos lingüísticos en un estudio sobre la lengua catalana.
6. KA31 (Conocimiento) Reconocer las distintas metodologías específicas para la recogida de datos lingüísticos en un estudio sobre la lengua catalana.
7. KA31 (Conocimiento) Reconocer las distintas metodologías específicas para la recogida de datos lingüísticos en un estudio sobre la lengua catalana.
8. KA32 (Conocimiento) Seleccionar las herramientas tecnológicas adecuadas para la gestión de corpus lingüísticos en un estudio sobre la lengua catalana.
9. SA31 (Habilidad) Realizar análisis estadísticos e interpretativos de distintas tipologías sobre datos lingüísticos de la lengua catalana en un estudio.
10. SA32 (Habilidad) Analizar datos lingüísticos de la lengua catalana con la ayuda de programas informáticos específicos.
11. SA33 (Habilidad) Explotar fuentes primarias lingüísticas con instrumentos digitales.
12. SA33 (Habilidad) Explotar fuentes primarias lingüísticas con instrumentos digitales.

Contenido

Bloque 1: Fundamentos para analizar la variación de una lengua

1. Fundamentos para analizar la variación interna de una lengua. Producción, percepción y subjetividad de los hablantes.
2. Preparación y ejecución de un proyecto de investigación: planteamiento de la investigación, marco conceptual, formulación de hipótesis (preguntas de investigación) y objetivos, utilidad de la investigación, determinación del diseño metodológico, cronograma.
3. Diseño metodológico: ubicación del trabajo, muestra, técnicas de recogida y análisis de datos (cuantitativos y cualitativos).
4. Introducción a la estadística: definición de variables, tipos y naturaleza de las variables, estadística descriptiva e inferencial, significación estadística.

Bloque 2: Tratamiento de datos lingüísticos con herramientas estadísticas

1. Introducción a R: preparación de datos: limpieza, transformación y codificación de variables lingüísticas.
2. Estadística descriptiva: resumen y visualización de datos: tablas, gráficos y medidas de tendencia central y dispersión.
3. Muestreo de distribuciones, estadística de muestras y parámetros poblacionales.

4. Contrastación de hipótesis y pruebas paramétricas: diseño de pruebas estadísticas para la variación lingüística.

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Asistencia a las clases y a las actividades programadas	24	0,96	CA23, CA24, KA30, KA31, KA32, SA31, SA32, SA33, CA23
Elaboración de trabajos y de actividades en casa	91	3,64	CA23, SA31, SA32, SA33, CA23

La asignatura combina sesiones teóricas con actividades prácticas y aplicadas. Los contenidos teóricos se complementarán con ejercicios en el aula, la manipulación de datos reales y la realización de prácticas con software estadístico. Los estudiantes desarrollarán una propuesta de investigación que integrará los conocimientos adquiridos a lo largo del curso y un informe de resultados realizado con el software R.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Ejercicio análisis estadístico	40%	3,75	0,15	KA32, SA31, SA32, SA33
Ejercicios prácticos	20%	2,5	0,1	CA23, KA31, KA32, SA31, SA32, SA33
Propuesta de un proyecto de investigación	40%	3,75	0,15	CA24, KA30, KA31

Ejercicios prácticos (20%): Se evaluará la realización regular y la calidad de los ejercicios prácticos propuestos durante las sesiones. Estos ejercicios estarán orientados al desarrollo de competencias metodológicas y técnicas relacionadas con el análisis de datos lingüísticos y el uso del software R. También se valorará la participación activa y el aprovechamiento de las actividades.

Propuesta de un proyecto de investigación (40%): Los estudiantes deberán elaborar una propuesta formal de proyecto de investigación en el ámbito de la variación lingüística.

Ejercicio de análisis estadístico (40%): El estudiante deberá entregar un ejercicio final que consista en el análisis estadístico de un conjunto de datos lingüísticos utilizando R. Este informe deberá incluir el tratamiento de los datos, la aplicación de técnicas estadísticas adecuadas, la interpretación de los resultados y la presentación clara y estructurada de las conclusiones. Se tendrá en cuenta tanto la corrección técnica como la capacidad de argumentación y exposición.

Bibliografía

- Butler, Christopher S. (1985). Statistics in Linguistics. Oxford, Basil Blackwell
- Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction*. Walter de Gruyter.
- Labov, W. (2010) Principles of Linguistic Change. Vol 3: Cognitive and Cultural Factors. Malden/Oxford: Wiley-Blackwell.
- Moreno-Fernández, F. (2012) Sociolingüística cognitiva. Proposiciones, escolios y debates. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Pradilla, M. À. (2008). Sociolingüística de la variació i llengua catalana. Barcelona: Institut d'Estudis Catalans
- Sonderegger, M. (2023). *Regression modeling for linguistic data*. MIT Press.
- Strelluf, C. (ed.) (2023) The Routledge Handbook of Sociophonetics. Londres: Routledge.
- Tagliamonte, S. (2012) Variationist Sociolinguistics. Change, Observation, Interpretation. Malden/Oxford: Wiley-Blackwell
- Tatham, M.; Morton, K. (2011). A guide to Speech Production and Perception. Edinburgh University Press: Edimburg.
- Verzani, John. (2005). Using R for introductory statistics. Boca Raton: Chapman & Hall.
- Vida-Castro, M.; Ávila-Muñoz, A. M. (eds.) (2024) The Continuity of Linguistic Change: Selected Papers in Honour of Juan Andreis Villena-Ponsoda. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Software

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.x.x) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>

Grupos e idiomas de la asignatura

La información proporcionada es provisional hasta el 30 de noviembre de 2025. A partir de esta fecha, podrá consultar el idioma de cada grupo a través de este [enlace](#). Para acceder a la información, será necesario introducir el CÓDIGO de la asignatura

Nombre	Grupo	Idioma	Semestre	Turno
(TEm) Teoría (máster)	1	Catalán	segundo cuatrimestre	tarde