



DATOS IDENTIFICATIVOS

Tratamento de Datos e Lingüística de Corpus

Materia	Tratamento de Datos e Lingüística de Corpus			
Código	V01M069V01101			
Titulación	Máster Universitario en Lingüística e as súas Aplicacións			
Descritores	Creditos ECTS 3	Sinale OB	Curso 1	Cuadrimestre 1c
Lingua de impartición	Castelán Galego			
Departamento				
Coordinador/a	Perez Guerra, Javier			
Profesorado	Perez Guerra, Javier			
Correo-e	jperez@uvigo.es			
Web	http://www.poslea.com			
Descrición xeral	Introdución á metodoloxía de investigación textual baseada en corpus de textos mediante o uso de ferramentas informáticas sinxelas que permitan a análise e extracción de información de corpus.			

Competencias de titulación

Código			
A1	El alumnado conocerá y será capaz de aplicar los métodos y las técnicas de análisis lingüístico		
A2	El alumnado conocerá las corrientes actuales de la lingüística		
A3	El alumnado conocerá las corrientes actuales de la lingüística aplicada		
A4	El alumnado será capaz de aplicar los conocimientos lingüísticos a la práctica		
A5	El alumnado adquirirá habilidades necesarias para analizar e investigar los mecanismos de los procesos de comunicación.		
A6	El alumnado dominará aspectos teóricos y prácticos de la investigación gramatical		
B1	El alumnado desarrollará habilidades cognitivas avanzadas (capacidad de síntesis, análisis y pensamiento crítico) relacionadas con el desarrollo y la creación de un perfil investigador		
B2	El alumnado será capaz de integrar conocimientos de campos de estudio diversos		
B3	El alumnado aprenderá a aprender, de forma autónoma, como base de su futura actividad investigadora		
B4	El alumnado será capaz de aplicar sus conocimientos y habilidades en la resolución de problemas en entornos nuevos o no familiares en contextos interdisciplinares		
B5	El alumnado adquirirá capacidad para concebir y diseñar proyectos de investigación con rigor académico		

Competencias de materia

Resultados previstos na materia	Tipoloxía	Resultados de Formación e Aprendizaxe
Familiarizar ao alumnado coas formas de proceder na investigación científica, tanto na súa dimensión teórica coma na práctica.	saber facer	A1 A2 A3 A6 B5
Proporcionar claves para un razonamento crítico sobre os métodos de investigación en lingüística e sobre o papel da linguaxe nos procesos de desenvolvemento e cambio social.	saber facer	A1 A5 B1

Contidos

Tema

1. Bloque teórico-práctico	- Introducción á lingüística de corpus: modelos teóricos. Tipos de córpora de textos. As anotacións e o seu tratamento automatizado. - Manexo de datos textuais: ferramentas informáticas. - Estudos clásicos de tipo lingüístico, literario e cultural feitos sobre córpora de textos.
2. Bloque práctico	- Creación de córpora electrónicos de textos: emprego de córpora xa existentes, elaboración de novos córpora (búsqueda de textos: escaneado, Internet, etc.). - Demostracións e exercicios de etiquetado: COCOA, labels, contadores, XML/SGML-TEI, etc. - Utilidades informáticas de manexo básico de córpora electrónicos: TACT, MonoConc, SARA, Wordsmith, UAM CorpusTool, BNC Baby (Xaira), BNC web, parsers en liña, etc.

Planificación

	Horas na aula	Horas fóra da aula	Horas totais
Actividades introdutorias	0.5	0	0.5
Sesión maxistral	11.4	18.8	30.2
Traballos tutelados	4	40.2	44.2
Observación sistemática	0	0.1	0.1

*Os datos que aparecen na táboa de planificación son de carácter orientador, considerando a heteroxeneidade do alumnado.

Metodoloxía docente

	Descrición
Actividades introdutorias	Presentación da asignatura (contidos, metodoloxía, avaliación). Presentacións dos elementos metodolóxicos informáticos cos que o alumnado debe estar familiarizado previamente.
Sesión maxistral	Exposicións teórico-prácticas sobre os contidos da asignatura.
Traballos tutelados	Exposición dun pequeno traballo de investigación no que o alumno/a faga uso de técnicas de lingüística de corpus aplicadas a casos ou proxectos de natureza lingüística.

Atención personalizada

Probas	Descrición
Observación sistemática	O desenvolvemento ordinario da materia non require atención personalizadas. De todos os xeitos, o profesorado estará dispoñible no seu horario habitual de titorías para calquera solicitude de atención personalizada.

Avaliación

	Descrición	Cualificación
Sesión maxistral	Os contidos desenvolvidos nas sesións teórico-prácticas serán avaliados a través da introdución metodolóxica obrigatoria do traballo final.	30
Traballos tutelados	Desenvolvemento documental e presentación oral dun pequeno proxecto de investigación lingüística que faga uso da metodoloxía de lingüística de corpus.	60
Observación sistemática	Participación no desenvolvemento docente da materia.	10

Outros comentarios sobre a Avaliación

A asistencia e participación regular (80%) nas sesións docentes presenciais son obrigatorias. Naqueles casos nos que a ausencia estea xustificada documentalmente, o alumnado poderá optar por unha avaliación final (100%) mediante o traballo tutelado. Esta forma de avaliación extraordinaria será tamén aplicada na segunda convocatoria.

Bibliografía. Fontes de información

Bibliografía recomendada

Barnbrook, G. (1996) *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.

Biber, D., S. Conrad e R. Reppen (1998) *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London: Longman.

McEnery, T. (2003) [Corpus linguistics]. Ed. R. Mitkov *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, pp.448-63.

McEnery, T. e A. Wilson (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Meyer, Ch. (2002) *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press. Stubbs, M. (1996) *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture*. Oxford: Blackwell.

Recomendacións
