

MÓDULO	MATERIA	CURSO	SEMESTRE	CRÉDITOS	TIPO
IV. Aplicaciones	2. Los corpus lingüísticos: aplicaciones	2019-2020	1º	3	Optativa
<b>PROFESORES*</b>			<b>DIRECCIÓN COMPLETA DE CONTACTO PARA TUTORÍAS (Dirección postal, teléfono, correo electrónico, etc.)</b>		
<ul style="list-style-type: none"> <li>Miguel Calderón Campos</li> <li>Gael Vaamonde Dos Santos</li> </ul>			Dirección postal: Dpto. de Lengua Española Campus de la Cartuja. Teléfono: 958249077 Correo electrónico: <a href="mailto:calderon@ugr.es">calderon@ugr.es</a> ; <a href="mailto:gaelvaamonde@ugr.es">gaelvaamonde@ugr.es</a>		
			Página electrónica: <a href="http://corpora.ugr.es/ode">http://corpora.ugr.es/ode</a>		
			<b>HORARIO DE TUTORÍAS*</b>  Consúltese en la página electrónica del Departamento de Lengua Española: <a href="http://www.lenguasp.ugr.es/profesorado">http://www.lenguasp.ugr.es/profesorado</a>		
<b>MÁSTER EN EL QUE SE IMPARTE</b>			<b>OTROS MÁSTERES A LOS QUE SE PODRÍA OFERTAR</b>		
Máster Oficial en Estudios Superiores de Lengua Española					
<b>PRERREQUISITOS Y/O RECOMENDACIONES (si procede)</b>					
<ul style="list-style-type: none"> <li>NIVEL B2 DE ESPAÑOL ACREDITADO. Muy aconsejado nivel C1.</li> <li>El carácter técnico de esta asignatura hace muy aconsejable haber cursado asignaturas del grado en Filología Española relacionadas con la lingüística hispánica, tanto sincrónica como diacrónica, de manera especial Historia de la lengua española</li> </ul>					
<b>BREVE DESCRIPCIÓN DE CONTENIDOS</b>					
Un corpus es un "conjunto de documentos lingüísticos de cualquier tipo, que se toma como modelo de un estado o nivel de lengua predeterminado, al cual representa o se pretende que represente" (Alvar Ezquerro y					

\* Consulte posible actualización en Acceso Identificado > Aplicaciones > Ordenación Docente.



otros 1994). En la actualidad, es difícil concebir un trabajo de investigación lingüística sin un corpus documental previo. Tenemos a nuestra disposición abundantes corpus en línea (CORDE, CREA, Corpus del español, Corpus del Nuevo Diccionario Histórico, CORDIAM, CORLEXIN (Corpus léxico de inventarios), corpus CHARTA, COREECOM (Corpus electrónico del español colonial mexicano), Post-Scriptum, etc. y herramientas informáticas para analizar corpus de diseño propio (WordSmith, AntConc, etc.).

El objetivo de este curso es mostrar las aplicaciones de los corpus digitales y de sus herramientas de análisis para la investigación lingüística de diversa naturaleza: estudios normativos, lexicográficos, descriptivos, de carácter sincrónico o diacrónico.

### **COMPETENCIAS GENERALES, TRANSVERSALES Y ESPECÍFICAS**

#### **GENERALES**

CG4 - Aplicar un razonamiento crítico a la solución de problemas.

CG6 - Gestionar información bibliográfica.

CG8 - Respetar la diversidad y la multiculturalidad.

CG9 - Trabajar en equipo.

CG10 - Aprender de forma autónoma.

CG12 - Afrontar problemas de investigación con una perspectiva interdisciplinar.

CB6 - 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

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB8 - Que los 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

CB9 - Que los 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

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

#### **TRANSVERSALES**

CT2 - Ser capaz de informarse sobre los cambios y nuevas tendencias del español en el mundo.

CT4 - Manejar con criterios profesionales las principales fuentes de referencia del español.

CT5 - Perfeccionar la escritura académica.

CT6 - Diseñar correctamente una investigación sobre temas lingüísticos. TRANSVERSALES

CT7 - Aplicar las Tecnologías de la Información y la Comunicación (TIC) al estudio y enseñanza de la lengua española.

CT10 - Compilar y manejar corpus lingüísticos al servicio de la investigación.

#### **ESPECÍFICAS**

CE1- Analizar los procesos de variación y cambio lingüístico.

CE2- Planificar futuras situaciones relacionadas con el español.

CE6- Aplicar distintos modelos teóricos al estudio del español.

CE10- Resolver problemas de transmisión textual y de edición de textos.

### **OBJETIVOS (EXPRESADOS COMO RESULTADOS ESPERABLES DE LA ENSEÑANZA)**



El alumnado sabrá:

- Aplicar los conocimientos necesarios para configurar el corpus base de su investigación lingüística.
- Formular hipótesis y preguntas de investigación sobre aspectos sincrónicos y diacrónicos a partir de la consulta o elaboración de corpus lingüísticos.
- Utilizar corpus lingüísticos y bases de datos para analizar fenómenos lingüísticos de naturaleza diacrónica o sincrónica.
- Buscar las fuentes bibliográficas y recursos para el análisis y la elaboración de corpus.
- Diseñar una posible investigación a partir de corpus relacionada con algún aspecto lingüístico de índole sincrónica o diacrónica.

### TEMARIO DETALLADO DE LA ASIGNATURA

TEMARIO TEÓRICO:

- Tema 1. Términos y conceptos básicos de la lingüística de corpus. XML-TEI. Tokenización, normalización, lematización y etiquetado morfosintáctico. Lenguaje CQL. Expresiones regulares.
- Tema 2. Tipos de corpus lingüísticos.
- Tema 3. Los corpus en línea: CORDE, CREA, Corpes XXI, Corpus del español, CorLexIn, CORDIAM, Sketch Engine, Post Scriptum, etc.
- Tema 4. Herramientas para la elaboración y diseño de corpus de investigación lingüística diacrónica y sincrónica: AntConc, LancsBox, TEITOK, etc.
- Tema 5. Análisis cualitativo y cuantitativo de datos a partir de corpus. Introducción a la estadística básica aplicada a los corpus lingüísticos.

TEMARIO PRÁCTICO:

Seminarios/Talleres

- Seminario 1. Lectura y análisis de artículos de investigación lingüística basados en análisis cualitativo o cuantitativo de corpus.
- Seminario 2. Microdiseños de investigaciones diacrónicas o sincrónicas basadas en corpus.
- Seminario 3. Prácticas de tokenización, normalización, etiquetado de corpus. Lenguaje CQL y expresiones regulares.

BIBLIOGRAFÍA FUNDAMENTAL:

- Bertolotti, Virginia, y Concepción Company (2014), "El corpus diacrónico y diatópico del español de América (CORDIAM). Propuesta de tipología documental", *Cuadernos de la ALFAL*, 6, 130-148.
- Clavería Nadal, Gloria (2012): "Corpus diacrónicos: nuevas perspectivas para el estudio de la historia de la lengua", en Emilio Montero Cartelle (ed.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española*, vol. I, Santiago de Compostela, Meubook, pp. 405-420.
- Vaamonde, Gael. 2015. "P.S. Post Scriptum. Dos corpus diacrónicos de escritura cotidiana. *Procesamiento del lenguaje natural*", 55, 57-64.
- Rojo, Guillermo (2010), "Sobre codificación y explotación de corpus textuales: otra comparación del Corpus del español con el CORDE y el CREA", *Lingüística*, 24, 11-50.
- Rojo, Guillermo (2011), "Frecuencia de inventario y frecuencia de uso", *Revista española de lingüística*, 41/1, 5-43.
- Sánchez-Prieto Borja, Pedro (2011), *La edición de textos españoles medievales y clásicos. Criterios de*



*presentación gráfica, San Millán de la Cogolla, Cilengua.*

#### **ENLACES RECOMENDADOS**

CREA <http://corpus.rae.es/creanet.html>

CORDE <http://corpus.rae.es/cordenet.html>

CORPUS DEL ESPAÑOL <http://www.corpusdelespanol.org/>

Programa de concordancias AntConc <http://www.laurenceanthony.net/software.htm>

Portal de Archivos Españoles. Pares <http://pares.mcu.es>

Corpus Post-Scriptum <http://ps.clul.ul.pt/index.php>

ODE (Oralia diacrónica del español): <http://corpora.ugr.es/ode>

CORDIAM: <http://www.cordiam.org/>

LancsBox: <http://corpora.lancs.ac.uk/lancsbox/>

#### **METODOLOGÍA DOCENTE**

- Teórica, argumentativa e indagadora (lección magistral).
- Analítica (comentarios de texto, lectura y crítica de artículos, previamente planificados, realizados individual o colectivamente).
- Analítica e indagadora (trabajos escritos de tipo crítico, con posible exposición y debate).
- Diálogo crítico entre profesor y alumno (tutorías individuales o en grupo).

#### **SISTEMAS DE EVALUACIÓN**

- La evaluación continua del aprendizaje estará basada en la realización de, al menos, dos pruebas a lo largo del curso: la primera, que supondrá el 30-40% de la calificación, se relacionará con los contenidos prácticos de la asignatura. La segunda prueba (60-70%) se centrará en los aspectos teórico-prácticos de la asignatura. Para garantizar la objetividad y en función del número de asistentes, podrán realizarse pruebas escritas, trabajos o exposiciones orales, que se concretarán con la suficiente antelación.

#### **INFORMACIÓN ADICIONAL**

