



Guía Docente				
Datos Identificativos				2015/16
Asignatura (*)	Tratamento de datos e lingüística de corpus	Código	613436001	
Titulación	Mestrado Universitario en Lingüística e as súas Aplicacións			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Primeiro	Obrigatoria	3
Idioma	CastelánGalegoInglés			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Filoloxía Inglesa			
Coordinación		Correo electrónico		
Profesorado		Correo electrónico		
Web	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 / Resultados do título	
Código	Competencias / Resultados do título
A3	O alumnado coñecerá as correntes actuais da lingüística aplicada.
A4	O alumnado será capaz de aplicar os coñecementos lingüísticos á práctica.
A5	O alumnado adquirirá habilidades necesarias para analizar e investigar os mecanismos dos procesos de comunicación.
B3	O alumnado aprenderá a aprender, de forma autónoma, como base da súa futura actividade investigadora.
B4	O alumnado será capaz de aplicar os seus coñecementos e habilidades na resolución de problemas en entornos novos ou non familiares en contextos interdisciplinares.
B6	O alumnado desenvolverá habilidades de xestión da información para propósitos académicos e de investigación (identificación e acceso a bibliografía, utilización apropiada de recursos e equipamentos, utilización de tecnoloxía para rexistrar datos, manexo de bases de datos, etc.).
B7	O alumnado adquirirá capacidade de organizar o traballo, planificar e xestionar o tempo e resolver problemas de forma efectiva.
B8	O alumnado será capaz de comunicar oralmente e por escrito con claridade os coñecementos teóricos e aplicados do campo de estudo tanto a audiencias expertas como non expertas.
C1	Expresarse correctamente, tanto de forma oral coma escrita, nas linguas oficiais da comunidade autónoma.
C3	Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.
C6	Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.
C7	Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.
C8	Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.

Resultados da aprendizaxe			
Resultados de aprendizaxe		Competencias / Resultados do título	
O alumnado será capaz de aplicar os coñecementos lingüísticos á práctica.		AI3	
		AI4	
O alumnado adquirirá habilidades necesarias para analizar e investigar os mecanismos dos procesos de comunicación.		AI5	BI3
O alumnado será capaz de aplicar os coñecementos lingüísticos á práctica.		AI4	
O alumnado será capaz de aplicar os seus coñecementos e habilidades na resolución de problemas en entornos novos ou non familiares en contextos interdisciplinares.			BI4



O alumnado desenvolverá habilidades de xestión da información para propósitos académicos e de investigación (identificación e acceso a bibliografía, utilización apropiada de recursos e equipamentos, utilización de tecnoloxía para rexistrar datos, manexo de bases de datos, etc.).		BI6	
O alumnado adquirirá capacidade de organizar o traballo, planificar e xestionar o tempo e resolver problemas de forma efectiva.		BI7	
O alumnado será capaz de comunicar oralmente e por escrito con claridade os coñecementos teóricos e aplicados do campo de estudo tanto a audiencias expertas como non expertas.		BI8	
Expresarse correctamente, tanto de forma oral coma escrita, nas linguas oficiais da comunidade autónoma.			CM1
Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.			CM3
Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.			CM6
Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.			CM7
Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.			CM8
Capacidade para localizar, utilizar, avaliar, sintetizar e aproveitar a bibliografía e outra información contida en bases de datos e outros instrumentos informáticos.			
Coñecemento dos métodos e técnicas de análise lingüística			
Coñecemento das correntes teóricas e metodolóxicas da lingüística			
Dominio instrumental avanzado oral e escrito da lingua inglesa			
Coñecemento da gramática da lingua inglesa			
Coñecemento da variación lingüística da lingua inglesa.			
Coñecemento da evolución histórica externa e interna da lingua inglesa.			
Capacidade para elaborar textos orais e escritos de diferente tipo en lingua inglesa.			
Capacidade para interrelacionar os distintos aspectos da Filoloxía.			
Capacidade para relacionar os coñecementos con outras áreas e disciplinas			
Aplicar un pensamento crítico, lóxico e creativo			
Capacidade de análise e síntese			
Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida			
Desenvolverse para o exercicio dunha cidadanía aberta, culta, crítica, comprometida, democrática e solidaria, capaz de analizar a realidade, diagnosticar problemas, formular e implantar solucións baseadas no coñecemento e orientadas ao ben común.			

Contidos	
Temas	Subtemas
1. La lingüística de corpus como metodología lingüística	1.1 Revisión histórica 1.2 Fundamentos de la lingüística de corpus 1.3 Lingüística de corpus vs. lingüística computacional
2. Concepto de corpus. Tipos.	2.1 Definiciones 2.2 Utilidad 2.3 Tipología
3. Diseño de un corpus:	3.1. Recopilación 3.2. Etiquetado. Normas y estándares 3.3. Explotación. Programas de concordancias, análisis cuantitativo vs. análisis cualitativo.
4. Algunos ejemplos.	4.1 Coruña Corpus 4.2 Corde, Crea 4.3 Helsinki Corpus, Lampeter Corpus

Planificación



Metodoloxías / probas	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas traballo autónomo	Horas totais
Proba de resposta breve		0	4	4
Actividades iniciais		1	0	1
Sesión maxistral		20	30	50
Lecturas		1	10	11
Prácticas a través de TIC		8	0	8
Atención personalizada		1	0	1

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

Metodoloxías	
Metodoloxías	Descrición
Proba de resposta breve	Realización de pequenos exercicios ao remate de cada tema.
Actividades iniciais	Presentación da asignatura (contidos, metodoloxía, avaliación). Presentación e introdución dos elementos metodolóxicos infpráticos cos que o alumnado debe estar familiarizado previamente.
Sesión maxistral	Exposicións teórico-prácticas sobre os contidos da materia.
Lecturas	Realización de lecturas sobre temas propostos que complementen as exposicións das sesións maxistrais.
Prácticas a través de TIC	Realizáranse traballos prácticos sobre creación e explotación de corpus empregando as TIC.

Atención personalizada	
Metodoloxías	Descrición
Actividades iniciais	Presentación da asignatura (contidos, metodoloxía, avaliación). Presentación e introdución dos elementos metodolóxicos infpráticos cos que o alumnado debe estar familiarizado previamente.

Avaliación			
Metodoloxías	Competencias / Resultados	Descrición	Cualificación
Proba de resposta breve		Exercicios que recapitulen os contidos de cada tema incluíndo as lecturas propostas.	30
Sesión maxistral		Resolución de problemas, debate e participación na aula.	10
Prácticas a través de TIC		Avaliarase un exercicio proposto relacionado coa explotación/creación de corpus.	60

Observacións avaliación

Fontes de información



Bibliografía básica

- AARTS, J. & MEIJS, W. (eds). 1986. *Corpus Linguistics II*, Amsterdam: Rodopi. AARTS, J., DE HANN, P. & OOSTDIJK, N. (eds.) 1993. *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi. BAKER, P. - HARDIE, A. - McENERY, T. 2006. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press (Glossaries in Linguistics). BARNBROOK, G. 1996. *Language and Computers: a practical introduction to the computer analysis of language*. Edinburgh University Press. BIBER, D. 1993. Representativeness in corpus design. *Literary and Linguistics Computing* 8/4: 243-57. BIBER, D. et al. 1994. ?The design and analysis of the ARCHER corpus: A progress report [A representative Corpus of Historical English Registers]?, in KYTÖ et al. (1994), pp. 3-6. BURNAGE, G. & DUNLOP, D. 1993. ?Encoding the British National Corpus?, in Aarts et al. (1993), pp. 79-95. CHURCH, K., Gale, W., Hanks, P. and Hindle, D. 1991. Using statistics in lexical analysis. In Zernik, U. (ed.) *Lexical Acquisition: Exploiting on-line resources to build a lexicon*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. DE HAAN, P. and van Hout, R. 1986. Statistics and corpus analysis, in Aarts, J., and Meijs, W. (eds.) *Corpus Linguistics II*, Amsterdam: Rodopi. EDWARDS, J.A. 1993. ?Survey of Electronic Corpora and Related Sources for Language Researchers?, in EDWARDS, J.A. & LAMPERT, M.D. pp. 263-310. GARCÍA MIGUEL, J.M. 2009. Enlaces de interés en Lingüística de Corpus. <
<http://webs.uvigo.es/web575/jm/m/enlaces-corpus.htm>>. GREENBAUM, S. 1992. ?A New Corpus of English: ICE?, in SVARTVIK, J. (1992), pp. 171-179. HALLIDAY, M. 1991. "Corpus studies and probabilistic grammar", in Aijmer and Altenberg 1991, pp 30-43. HALTEREN, H. & OOSTDIJK, N. 1993. ?Towards a syntactic Database: the TOSCA analysis system?, in AARTS et al, pp. 145-161. HICKEY, R. 1994. ?Applications of software in the compilation of corpora?, in KYTÖ et al. 165-186. HOLMES, J. .1994 "Inferring language change from computer corpora: some methodological problems", *ICAME Journal* 18: 27-40. JIMÉNEZ RUIZ, J.L. 2003. Los productos lingüísticos de la actividad epistemológica teórica de la lingüística objetual, *ELUA*, 17: 387-405. JOHANSSON, S. & STENSTRÖM, A. (eds.) 1991. *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. JOHANSSON, S. 1991. ?Computer Corpora in English Language Research?, in JOHANSSON, S. & STENSTRÖM, A. (1991), pp. 3-6. JOHANSSON, S. 1994. ?Some aspects of the recommendations of the Texts Encoding Initiative, with special reference to the encoding of language corpora?, in KYTÖ et al. (1994), pp. 203-212. JOHANSSON, S. 1995. ?Mens Sana in Corpore Sano: On the Role of Corpora in Linguistic Research?, *The European English Messenger*, IV/2, 19-25. KENNEDY, G. (1992) "Preferred ways of putting things", in Svartvik J. (ed) *Directions in Corpus Linguistics*, pp 335-73, Berlin: Mouton de Gruyter. KUCERA, H. 1992. ?The Odd Couple: The linguist and the Software Engineer. The Struggle for High Quality Computerized Language Aids?, in SVARTVIK, J. (1992), pp. 401-420. KYTÖ, M. 1993. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and List of Source Texts*. Helsinki: Helsinki University Press. KYTÖ, M. RISSANEN, M. & WRIGHT, S. (eds.) 1994. *Corpora Across the Centuries. Proceedings of the First International Colloquium on English Diachronic Corpora*. Amsterdam: Rodopi. LAVID, J. 2005. *Lenguaje y nuevas tecnologías nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Cátedra. LEECH, G. & GARSIDE, R. 1991. ?Running a Grammar Factory: The Production of Syntactically Analysed Corpora or ?Treebanks??, in JOHANSSON, S. & STENSTRÖM, A. (1991), pp. 15-32. LEECH, G. 1992 "Corpora and theories of linguistic performance", in Svartvik, J. *Directions in Corpus Linguistics*, pp 105-22. Berlin: Mouton de Gruyter. LEECH, G. 1993. ?Corpus Annotation Schemes?. *Literary and Linguistic Computing*, 8/4: 275-281. LEZCANO, E., PÉREZ GUERRA, J. & SEOANE POSSE, E. ?English Corpus Linguistics and Historical Research?. In GONZÁLEZ Y FERNÁNDEZ CORUGEDO, S. (ed.) *Some sundry wits gathered together*. Universidade da Coruña. (forthcoming). MARKUS, M. 1994. ?The Concept of ICAMET (Innsbruck Computer Archive of Middle English Texts)?, in KYTÖ et al. 41-52. MARTÍ ANTONÍN M.A. & castellón Masalles, I. 2000. *Lingüística computacional*. barcelona: Universitat de Barcelona. McENERY, A., Baker, P. and Wilson, A. 1995 "A statistical analysis of corpus based computer vs traditional human teaching methods of part of speech analysis.", *Computer Assisted Language Learning* 8(2/3): 259-74. McENERY, T & WILSON, A. 1996. *Corpus Linguistics*. Edinburgh University Press. PÉREZ HERNÁNDEZ, Chantal. 2002. *Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. <
<http://elies.rediris.es/elies18/>>. QUIRK, R. 1992. ?On Corpus principles and Design?, in SVARTVIK, J. 1992, pp. 457-469. REPPEN, R., Susan M. FITZMAURICE and Douglas BIBER. eds. 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam/New York: John Benjamins. RISSANEN, M. 1989. ?Three Problems Connected with the Use of Diachronic Corpora?, *ICAME Journal*, 13: 16-19. RISSANEN, M. 1994. ?The Helsinki Corpus of English Texts?, in



KYTÖ et al. (1994), pp. 73-79. SAMPSON, G. 1992. "Probabilistic parsing", in Svartvik, J. *Directions in Corpus Linguistics*, pp. 425-47. Berlin: Mouton de Gruyter. SAMPSON, G. 1992. "Analysed Corpora of English: A Consumer Guide?", in PENNINGTON, M.C. & STEVENS, V. (1992), pp. 181-200. SANTALLA DEL RÍO, M.P. 2005. La elaboración de corpus lingüísticos. En Cal, M., Núñez, P. & Palacios, I (eds.) *Nuevas tecnologías en lingüística, traducción y enseñanza de lenguas*. Santiago de Compostela: Universidade de Santiago de Compostela [SINCLAIR, J.M. 1992. "The Automatic Analysis of Corpora?", in SVARTVIK, J. (1992), pp. 379-397. SOUTER, C. 1993. "Towards a standard format for parsed corpora?", in AARTS et al. (1993), pp. 197-212. SPERBERG-MCQUEEN, C. M. and BURNARD, L. (1994) *Guidelines for Electronic Text Encoding and Interchange (P3)*. Chicago and Oxford: Text Encoding Initiative. STUBBS, M. 2001. *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell. SVARTVIK, J. 1992. "Corpus Linguistics comes of Age?", in SVARTVIK, J. (1992), pp. 7-13. SVARTVIK, J. 1992. *Directions in Corpus Linguistics*. Berlin: Mouton. TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at Work. Studies in Corpus Linguistics 6* TORRUELLA, J. & LLISTERRI, J. 1999: "Diseño de corpus textuales y orales", en J. M. Blecua, G. Clavería, C. Sánchez y J. Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio, pp. 45-77. WOODS, A., FLETCHER, P., & HUGHES, A. 1986. *Statistics in Language Studies*. Cambridge. Cambridge University Press.



Bibliografía complementaria

Recomendacións

Materias que se recomenda ter cursado previamente

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións

Recomendase ter un coñecemento básico de lingua inglesa.

(*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías