



## Teaching Guide

Identifying Data				2015/16	
<b>Subject (*)</b>	Tratamento de datos e lingüística de corpus	<b>Code</b>	613436001		
<b>Study programme</b>	Mestrado Universitario en Lingüística e as súas Aplicacións				
Descriptors					
<b>Cycle</b>	<b>Period</b>	<b>Year</b>	<b>Type</b>	<b>Credits</b>	
Official Master's Degree	1st four-month period	First	Obligatoria	3	
<b>Language</b>	SpanishGalicianEnglish				
<b>Teaching method</b>	Face-to-face				
<b>Prerequisites</b>					
<b>Department</b>	Filoloxía Inglesa				
<b>Coordinador</b>		<b>E-mail</b>			
<b>Lecturers</b>		<b>E-mail</b>			
<b>Web</b>	www.poslea.com				
<b>General description</b>	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.				

## Study programme competences

Code	Study programme competences
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.

## Learning outcomes

Learning outcomes	Study programme competences		
O alumnado será capaz de aplicar os coñecementos lingüísticos á práctica.	AR3		
	AR4		
O alumnado adquirirá habilidades necesarias para analizar e investigar os mecanismos dos procesos de comunicación.	AR5	BR3	
O alumnado será capaz de aplicar os coñecementos lingüísticos á práctica.	AR4		
O alumnado será capaz de aplicar os seus coñecementos e habilidades na resolución de problemas en entornos novos ou non familiares en contextos interdisciplinares.		BR4	



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.).		BR6	
O alumnado adquirirá capacidade de organizar o traballo, planificar e xestionar o tempo e resolver problemas de forma efectiva.		BR7	
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.		BR8	
Expresarse correctamente, tanto de forma oral coma escrita, nas linguas oficiais da comunidade autónoma.			CC1
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.			CC3
Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.			CC6
Asumir como profesional e cidadán a importancia da aprendizaxe ao longo da vida.			CC7
Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.			CC8
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.			

Contents	
Topic	Sub-topic
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

<b>Planning</b>
-----------------



Methodologies / tests	Competencies	Ordinary class hours	Student's personal work hours	Total hours
Short answer questions		0	4	4
Introductory activities		1	0	1
Guest lecture / keynote speech		20	30	50
Workbook		1	10	11
ICT practicals		8	0	8
Personalized attention		1	0	1

(\*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies	
Methodologies	Description
Short answer questions	Realización de pequenos exercicios ao remate de cada tema.
Introductory activities	Presentación da asignatura (contidos, metodoloxía, avaliación). Presentación e introdución dos elementos metodolóxicos informáticos cos que o alumnado debe estar familiarizado previamente.
Guest lecture / keynote speech	Exposicións teórico-prácticas sobre os contidos da materia.
Workbook	Realización de lecturas sobre temas propostos que complementen as exposicións das sesións maxistras.
ICT practicals	Realizáranse traballos prácticos sobre creación e explotación de corpus empregando as TIC.

Personalized attention	
Methodologies	Description
Introductory activities	Presentación da asignatura (contidos, metodoloxía, avaliación). Presentación e introdución dos elementos metodolóxicos informáticos cos que o alumnado debe estar familiarizado previamente.

Assessment			
Methodologies	Competencies	Description	Qualification
Short answer questions		Exercicios que recapitulen os contidos de cada tema incluíndo as lecturas propostas.	30
Guest lecture / keynote speech		Resolución de problemas, debate e participación na aula.	10
ICT practicals		Avaliarase un exercicio proposto relacionado coa explotación/creación de corpus.	60

Assessment comments

Sources of information



## Basic

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.



Complementary	
---------------	--

### Recommendations

Subjects that it is recommended to have taken before

Subjects that are recommended to be taken simultaneously

Subjects that continue the syllabus

Other comments

Recomendase ter un coñecemento básico de lingua inglesa.

(\*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.