



Universidade de Vigo

# Lingüística informática e de corpus aplicadas á lingua inglesa

*(English Corpus Linguistics)*

Zeltia Blanco Suárez (USC)

Yolanda Fernández Pena (USC)

Isabel Moskowich-Spiegel (UDC), coord.

David Tizón Couto (UVigo)

Nuria Yáñez Bouza (UVigo)

**GUÍA DOCENTE E MATERIAL DIDÁCTICO**

**2022/2023**

***MÁSTER INTERUNIVERSITARIO EN ESTUDOS INGLESES AVANZADOS E AS SÚAS APLICACIÓNS***

AUTORES: Zeltia Blanco Suárez, Yolanda Fernández Pena, Isabel Moskowich-Spiegel, David Tizón Couto, Nuria Yáñez Bouza

Edición electrónica. 2022

ADVERTENCIA LEGAL: Reservados todos os dereitos. Queda prohibida a duplicación total ou parcial desta obra, en calquera forma ou por calquera medio (electrónico, mecánico, gravación, fotocopia ou outros) sen consentimento expreso por escrito dos autores.

**COURSE TITLE**

English corpus linguistics

**ECTS CREDITS**

3 ECTS credits

**COURSE TYPE**

Optional

**SEMESTER**

2nd semester (Spring)

**PREREQUISITES**

None

**AIMS AND SKILLS**

The aim of this course is to introduce the field of corpus linguistics and corpus-based methodology for the linguistic study of English. General issues about the design, collection and analysis of corpus material will be discussed. In practical sessions students will learn how to search and analyse text corpora using software tools. Students will develop the ability to describe and discuss the extent to which language usage varies and can be tracked with reference to genuine examples attested in corpora.

With this course students will acquire skills in:

- the principles and methodology of English corpus linguistics;
- the practical use of linguistic corpora and related software;
- the discussion of specialized literature on corpus linguistics;
- the presentation of research results.

## COMPETENCES

G01 - Ability to delve into those concepts, principles, theories or models related to the various areas of English Studies, as well as to become familiar with the methodology required to solve those problems typical of this field of study.

G02 - Ability to apply the knowledge gained/obtained within the multidisciplinary and multifaceted/versatile area of English Studies.

G03 - Ability to use the new information and communication technologies (ICTs) efficiently within the realm of English Studies.

G04 - Ability to present experiences, ideas or reports in public, as well as to express informed opinions based on criteria, external rules or personal reflections, for which a sufficient command of the academic and scientific language, both written and oral, will be necessary.

G05 - Abilities to investigate and manage new knowledge and information within the context of English Studies.

G06 - Ability to acquire/achieve critical thinking that will lead students to consider the relevance of the existing research in the fields of study that make up/shape/define English Studies, as well as the relevance of their own investigations.

G07 - Abilities to consolidate and develop the C2 linguistic competence in the spoken and written use of the English language.

G08 - Progressive autonomy in the learning process, personal/individual search for resources and information, by getting access/accessing (to) bibliographical and documentary references on the various areas that make up/define/shape English Studies.

G09 - Ability to carry out research work of an academic nature in the different realms of English Studies.

E01 - Knowledge of the main methods of linguistic research.

E02 - Knowledge of the main resources, tools and methods for linguistic research.

E14 - To know how to apply techniques and methods typical of quantitative linguistic analysis.

## CONTENTS

- Introduction (description vs. theory; corpus vs. computational linguistics; brief history of corpus linguistics).
- What is a corpus? (defining a corpus, types of corpora, corpus resources).
- Corpus design and compilation (size, representativeness, etc.).
- Corpus annotation (tagging, parsing, other types of annotation).
- Data retrieval (software tools, concordances, wildcards, keywords, word lists, etc.).
- Data analysis (quantitative and qualitative analyses, normalized frequencies, frequency distribution, statistical significance).
- Applications of corpora for the linguistic analysis of English.
- Hands-on demonstrations with a selection of corpora.

## TEACHING METHODS

- Lectures on the role of corpus linguistics as a framework/methodology for linguistic investigation.
- Seminars and hands-on sessions with software tools and corpora for text analysis (available at computer clusters and via internet).
- Oral presentations of the students' case studies.
- eLearning platform for summative and formative assessment.
- Handouts will be provided to accompany lectures and seminars.

<b>ACADEMIC ACTIVITES</b>
---------------------------

<u>Activity</u>	<u>Hours</u>	<u>% Attendance</u>
Lectures	5h	100%
In-class interactive seminars	7h	100%
Tutoring	2h	100%
Use of virtual teaching platform (including reading activities, self-evaluation activities)	10h	0%
Assignments and oral presentation	31h	0%
Recommended readings/bibliographic search	15h	0%

<b>ASSESSMENT</b>
-------------------

**First opportunity:**

- 15% Active participation in the sessions. Please note that attendance is compulsory.
- 55% Exercises and set readings.
- 30% Oral presentation of a case study. Students will have to provide a hand-out and a PowerPoint presentation.

Exercises as well as the hand-out and the PowerPoint presentation for the case study will be submitted via the teaching platform.

**Second opportunity:**

Students will have to repeat only those parts (exercises and case study) in which they did not reach a pass in the first opportunity. In the second opportunity the case study will be submitted as a written essay (between 2,000-3,000 words, excluding references).

**Students officially exempt from class attendance:**

Students will have to hand in all course exercises (60%) and submit an essay on a case study (between 2,000-3,000 words, excluding references) (40%).

All exercises and tests will be done in English. Overall correctness in language and in the use of formal conventions is essential. Therefore, grammatical and orthographic mistakes will be penalised.

## RECOMMENDED BIBLIOGRAPHY

- Aarts, Jan, Pieter de Haan & Nelleke Oostdijk (eds.). 1993. *English language corpora: Design, analysis and exploitation*. Amsterdam: Rodopi.
- Baker, Paul. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baker, Paul, Andrew Hardie & Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, Douglas & Randi Reppen (eds.). 2015. *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics. A practical guide*. Cambridge University Press.
- Cantos, Pascual. 2011. *Statistical methods in language and linguistic research*. London: Equinox.
- Grant, Tim, Urszula Clark, Gertrud Reershemius, David Pollard, Sarah Hayes & Plappert Garry. 2017. *Quantitative research methods for linguists: A questions and answers approach for students*. New York: Routledge.
- Hoffmann, Sebastian, Stefan Evert, Nick Smith, David Lee & Ylva Berglund Prytz. 2008. *Corpus linguistics with BNCweb - a practical guide*. Frankfurt am Main: Peter Lang.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. London: Longman.
- Kilgariff, Adam & Gregory Grefenstette. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics* 29(3). 333-347.
- Lavid, Julia. 2005. *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Cátedra.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.
- Lindquist, Hans. 2009. *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
- Lüdeling, Anke & Merja Kytö (eds.). 2008. *Corpus linguistics. An international handbook*. Volume I. Berlin/New York: Walter de Gruyter.
- McEnery, Tony & Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies. An advanced resource book*. London: Routledge.
- Meyer, Charles. 2002. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Mitkov, Ruslan (ed.). 2003. *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.
- Mukherjee, Joybrato. 2004. The state of the art in corpus linguistics: Three book-length perspectives. *English Language and Linguistics* 8(1). 103-119.

Oakes, Michael. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.  
Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.  
Stubbs, Michael. 1996. *Text and corpus analysis*. Oxford: Blackwell.

### **ADDITIONAL INFORMATION**

Students are expected to complete all assignments and readings suggested by the lecturers, and to come to the sessions prepared to discuss them. Students must visit the eLearning platform for the course regularly. Students are expected to check their university email on a regular basis: announcements and last-minute changes will be notified via the teaching platform.

Academic misconduct (including cheating, plagiarism, self-plagiarism, collusion or fabrication of results) will not be tolerated and will be penalised.

- Zeltia Blanco Suárez (zeltia.blanco@usc.es)
- Yolanda Fernández Pena (yolanda.fernandez.pena@usc.es)
- Isabel Moskowich-Spiegel (imoskowich@udc.es)
- David Tizón Couto (davidtizon@uvigo.es)
- Nuria Yáñez Bouza (nuria.y.b@uvigo.es)