

Module/Course Title: Corpus Processing

- **Code number:**
- **Level of Module/Course (under-/postgraduate):**
- **Type of Module/Course:**
- **Year of Study**
- **Semester**
- **Number of ects allocated:**
- **Number of teaching units:**
- **Name of lecturer / lecturers :**
- **Content outline:**

A Corpus is an organized, statistically correct and constructed using specific criteria "collection" of real language material in electronic form, representative of a language, dialect, language for specific purposes, language variety, sublanguage. Corpus-based studies deal with phenomena of real language use. Corpora play a very important role in various linguistics areas research such as phonology, morphology, syntax, semantics, pragmatics, discourse analysis, lexicography, terminology, text linguistics, stylistics, sociolinguistics, historical linguistics, applied linguistics, dialectology, forensic linguistics and more. Corpus Processing gives precision, completeness and speed to quantitative as well as qualitative analysis of the language in a way that could not be achieved using traditional techniques. Nowadays, besides Linguistics, applications of corpus processing include a great variety of disciplines that involve the study of language: Education - First and 2nd/Foreign Language Learning and Teaching, Philology, Political Sciences, Media Studies, Psychology, Sociology and more.

During the course, students are introduced to:

the advantages of the applications of corpus processing to linguistics as well as other scientific disciplines that involve the study of language,

corpora categories and types,

corpus construction,

frequency normalization,

corpus annotation.

Emphasis is given to application, where students are introduced to the Linux environment and learn how to program for the extraction of basic, important products of corpus processing.

- **Learning outcomes:**

After the successful complete of the course, students should:

- Know what a Corpus is.
- Know the applications of Corpus Linguistics.
- Know the need for Corpus Processing to Linguistics.
- Know Corpus Processing applications to various disciplines besides Linguistics.
- Know the categorizations of corpora.
- Know about corpus construction.
- Know what normalization is.
- Know Greek Corpora available on the internet.
- Know about the Linux environment and its use for Corpus Processing.
- Know the basics of the Linux environment.
- Know basic Linux commands.
- Be able to program in the Linux environment in order to extract word frequency lists from corpora.

The course *Corpus Processing* gives students with a theoretical background the ability to deal with issues and problems in a formal, algorithmic way, an ability acquired with the study of mathematics/computing subjects.

Moreover, the students are introduced to the first steps of computer programming.

- **Prerequisites:** -

- **Recommended Reading:**

a) Basic Textbooks:

Φραντζή, Κ. 2012. Εισαγωγή στην Επεξεργασία Σωμάτων Κειμένων. Αθήνα: Ίων.

Γούτσος, Διονύσης και Γεωργία Φραγκάκη 2015. Εισαγωγή στη Γλωσσολογία Σωμάτων Κειμένων. Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - Αποθετήριο "Κάλλιπος".

Μικρός, Γεώργιος 2009. Ποσοτική Ανάλυση της Κοινωνιογλωσσολογικής Ποικιλίας - Θεωρητικές και Μεθοδολογικές Προσεγγίσεις. Αθήνα : Μεταίχμιο.

b) Additional References:

Baker, Paul (2009) *Contemporary Corpus Linguistics* (Contemporary Studies in Linguistics). Continuum International Publishing Group Ltd.

Biber, Douglas, Conrad, Susan and Randi Reppen, 1998. *Corpus linguistics: investigating language structure and use*. Cambridge University Press.

Cheng, Winnie (2011) *Exploring Corpus Linguistics: Language in Action* (Routledge Introductions to Applied Linguistics). Routledge.

Hyland, Ken (2013) *Corpus Applications in Applied Linguistics*. Bloomsbury Academic.

Kennedy, Graeme (1998) *An Introduction to Corpus Linguistics*. Longman.

McEnery, Anthony, Xiao, Richard and Yukio Tono, 2006. *Corpus-based Language Studies: an advanced resource book*. Routledge, 2006.

McEnery, Anthony (2011) *Corpus Linguistics: Method, Theory and Practice* (Cambridge Textbooks in Linguistics). Cambridge University Press

O'Keeffe, Anne (2012) The Routledge handbook of corpus linguistics (Routledge Handbooks in Applied Linguistics). Routledge.

Stubbs Michael (1996) Text and Corpus Analysis: Computer Assisted Studies of Language and Institutions (Language in Society). John Wiley & Sons.

Teubert Wolfgang (2007) Corpus Linguistics: A Short Introduction. Continuum.

- **Learning Activities and Teaching Methods:** interactive lectures
- **Assessment/Grading Methods:**
 1. end-of-semester exam
 2. assignment
- **Language of Instruction:** Greek/English
- **Mode of delivery (face-to-face, distance learning):** face-to-face