COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
Corpus Linguistics and Phraseology (Tekstynų lingvistika ir frazeologija)	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: Assoc. Prof. Dr Rita Juknevičienė **Other(s):**	Faculty of Philology 5 Universiteto St., LT-01513 Vilnius Tel. +370 5 268 7207 E-mail flf@flf.vu.lt

Study cycle	Type of the course unit (module)
MA (Magistro studijų studijų pakopa)	Compulsory (for English Studies (Linguistics) Elective (for English Studies (other tracks) (Privalomasis Anglistikos (kalbotyros krypties) magistro programoje; Pasirenkamasis kitų Anglistikos magistro programos krypčių studentams)

Mode of delivery	Course unit delivery period	Language(s) of instruction
Lectures, seminars (Kontaktinės paskaitos ir seminarai)	Spring Term	English

Requirements for students	
Prerequisites: Courses in linguistics, lexicology, corpus linguistics	**Additional requirements (if any):** English (C1)

Number of ECTS credits allocated	Total student's workload	Contact hours	Self-study hours
10	260	32	228

Purpose of the course unit (module): programme competences to be developed

The purpose of the course unit is to deepen students' understanding of corpus-driven and corpus-based linguistic research, with a particular focus on linguistic studies of (semi-)fixed multi-word expressions and formulaic language in English. The students will develop the following competences:

I. generic competences:
- cognitive;
- social;
- personal;

II. subject-specific competences:
- knowledge of contemporary linguistic theories and methods and their application to the study of the English language;
- research competence.

Learning outcomes of the course unit (module)	Teaching and learning methods	Assessment methods

Learning outcomes from the generic competences: - ability to find, analyse, synthesise and evaluate data needed for the studies; - ability to apply the acquired knowledge in practice, identify problems and propose solutions; - ability to generate ideas and choose appropriate forms of expression; - ability seek new knowledge and skills and apply them for tasks in a new environment; - ability to recognize different cultural identities and attitudes, explore others' ideas despite cultural differences, express one's ideas in ways that are understandable to others; - ability to take responsibility for the quality of one's own and the team's activities; - ability to take initiative and carry out one's activities, rationally assess the strengths and weaknesses of one's work, reflect and be responsible for one's decisions and actions. Learning outcomes from the subject-specific competences: - ability to effectively communicate in English in writing and orally; analyse and produce academic texts in English following relevant conventions and genre requirements; - ability to make effective use of modern information and communication technologies, databases and resources (corpora, corpus processing tools, dictionaries); - ability to raise a relevant research question in linguistics and design a research study in the field of corpus-based or corpus-driven phraseology; - ability to select the theoretical framework(s), critically evaluate relevant research methods or approaches and empirical material; - ability to conduct a research study by adopting corpus-based and corpus driven approaches to data collection and processing, interpreting findings of the study, drawing conclusions and evaluating the results in the context of other studies; - ability to present the results of individual research projects in writing and orally to a variety of audiences and demonstrate the ability to participate in discussions.	Interactive lectures, discussions, seminars, individual study, hands-on experience with corpora, mini-research projects	Continuous assessment (50%): a mini-research project for the seminars (25%) and a midterm test (25%). Written examination (50%): a written test consisting of open-ended questions.

Content: breakdown of the topics	Contact hours	Self-study work: time and assignments

	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work placement	**Contact hours**	**Self-study hours**	**Assignments**
1. Introduction to the course unit: aims, structure, assessment.	1						**1**		
2. Introduction to Corpus Linguistics. Key terms and issues in Corpus Linguistics. Corpora in different branches of linguistics (genre analysis, semantics, lexicography, language teaching, translation etc.).	2						**2**	**26**	Reading selected chapters from Tognini-Bonelli (2004); Stubbs (2007); Widdowson (2000)
3. Corpus compilation and design criteria: representativeness, balance and sampling, size. Different types of corpora. WEB as a corpus. English corpora.	2						**2**	**8**	Reading selected chapters from Gatto (2014); Sinclair (1991, 2004); Biber and Conrad (2009); Brezina (2018)
4. Making statistical claims: introduction to statistics in corpus linguistics (raw vs relative frequency; dispersion; statistical significance; chi-square tests; introduction to inferential statistics).	2		2				**4**	**32**	REading selected chapters from Brezina (2018)
5. Using corpora for language teaching research: learner corpus research; developing teaching materials; second language acquisition.	2						**2**	**22**	Reading selected chapters from McEnery et al. (2006); O'Keeffe et al. (2007); Granger (1998)
6. Corpus-based and corpus-driven phraseological research: overview of the field (idioms, collocations, lexical bundles/n-grams). Statistical tools in the identification of phrases.	7		7				**14**	**40**	Reading selected chapters: Gries (2008); Juknevičienė (2017); Biber et al. (2014); practical tasks on data retrieval and analysis.
7. Research tools and corpus analysis software: AntConc, LancsBox, SketchEngine, Wordsmith Tools.			2				**2**	**26**	Tasks on data retrieval and analysis
8. Practical exercises in replicating studies of corpus-driven research on English and translation to/from English.			4				**4**	**42**	Tasks and presentations of individual projects
9. Course summary and assessment.			1				**1**	**32**	
Total	**16**		**16**				**32**	**228**	

Assessment strategy	**Weight,%**	**Deadline**	**Assessment criteria**
Continuous assessment			
A mid-term test on theory	25	mid-term test	Quality of the answers to five open-ended questions; appropriate use of linguistic terminology and linguistic accuracy; specific requirements will be presented during the course. Test format: paper-and-pencil in-class written test.
A presentation of an individual research project	25	end of term	Content (clarity and appropriateness of information presented; conformity to presentation requirements;

			specific requirements will be presented during the course); appropriateness of methodological choices; awareness of project limitations.
Examination			
A written test	50	examination date	Format and focus: a written test consisting of 10 open-ended questions, involving theoretical and practical aspects of corpus linguistics. Assessment criteria: quality of the answers; appropriate use of linguistic terminology and linguistic accuracy.
The use of GenAI for study purposes should follow the general principles recommended by the Faculty of Philology. The students are encouraged to use GenAI for learning purposes and exploration of the subject matter, but they may not use GenAI to generate data necessary for individual projects or for the preparation of presentation slides.			

Author	**Year of publication**	**Title**	**Issue of a periodical or vol. of a publication**	**Publishing place and house or web link**
Compulsory reading				
Biber, D. & S. Conrad	2009	Register, Genre and Style		Cambridge University Press
Biber, D., S. Conrad & V. Cortes	2004	*If you look at…:* Lexical bundles in university teaching and textbooks	25/3: 371–405	*Applied Linguistics* (Oxford Journals Online)
Brezina, V.	2018	*Statistics in Corpus Linguistics*		Cambridge University Press
Crawford, W. J. & E. Csomay	2016	*Doing Corpus Linguistics*		Routledge / Taylor and Francis Group
Granger, S. & F. Meunier (eds.)	2008	*Phraseology: An interdisciplinary perspective*		Amsterdam: John Benjamins
Gries, S. Th	2008	Phraseology and linguistic theory: A brief survey.	3–26	Amsterdam: J. Benjamins Publishing Company. (In S. Granger & F. Meunier (Eds) *Phraseology. An Interdisciplinary Perspective*)
Juknevičienė, R.	2017	*English Phraseology and Corpora: An introduction to corpus-based and corpus-driven phraseology*		Vilnius: Vilniaus universiteto leidykla
McEnery, M. , R. Xiao & Y. Tono	2006	*Corpus-Based Language Studies*		London/New York: Routledge
Sinclair, J.	1991	*Corpus, Concordance, Collocation*		Oxford: Oxford University Press
Sinclair, J.	1996	The Search for Units of Meaning.	IX (1): 76–106	*TEXTUS*
Stubbs, M.	2007	Notes on the History of Corpus Linguistics and Empirical Semantics	317–329	Joensuu: Joensuun Yliopisto. (In M. Nenones & S. Niemi (Eds) *Collocations and Idioms*)
Tognini-Bonelli, E.	2004	*Corpus Linguistics at Work*		Amsterdam: John Benjamins
Optional reading				
Aijmer, K. (ed).	2009	*Corpora and Language Teaching*		Amsterdam/Philadelphia: John Benjamins
Conklin, K. & N. Schmitt	2008	Formulaic sequences: are they processsed more quickly than nonformulaic language by native and nonnative speakers?	29/1: 72–89	*Applied Linguistics* (Oxford Journals Online)
Cortes, V.	2015	Situating lexical bundles in the formulaic language spectrum	197–216	Amsterdam: J. Benjamins Publishing Company. (In V. Cortes & E. Csomay (Eds) *Corpus-based*

				Research in Applied Linguistics: Studies in Honor of Doug Biber)
Gatto, M.	2014	*Web as Corpus. Theory and Practice*		London/New Delhi/ New York/ Sydney: Bloomsbury
Granger, S. (Ed.)	1998	*Learner English on Computer*		London: Longman
Granger, S.	2019	*Formulaic Sequences in Learner Corpora: Collocations and Lexical Bundles*	228–247	New York: Routledge (In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds) *Understanding Formulaic Language*
Hyland, K	2008	As can be seen: Lexical bundles and disciplinary variation	27: 4–10	*English for Specific Purposes* (Elsevier)
Meunier, F. & S. Granger (Eds.)	2008	*Phraseology in Foreign Language Learning and Teaching*		Amsterdam: John Benjamins Publishing Company
McEnery, T. & A. Wilson	2001	*Corpus Linguistics*		Edinburgh University Press
Nesselhauf, N.	2005	*Collocations in a Learner Corpus*		Amsterdam: John Benjamins
O'Keeffe, A., McCarthy, M. & R. Carter	2007	*From Corpus to Classroom*		Cambridge: Cambridge University Press
Sinclair, J	2004	*Trust the Text*		London: Routledge
Stubbs, M.	2001	Texts, Corpora and Problems of Interpretation: A Response to Widdowson	22(2):149–172	*Applied Linguistics*
Vilkaite, L.	2016	Formulaic language is not all the same: comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs	8: 28–54	*Taikomoji kalbotyra*
Widdowson, H. G.	2000	On the Limitations of Linguistics Applied	21(1):3–25	*Applied Linguistics*
Wray, A.	2008	Formulaic Language: Pushing the Boundaries	57/2: 182–197	*ELT Journal* (Oxford Journals Online)
Corpus tools				
Anthony, L.	2018	AntConc [Computer Software]	v. 3.5.7	Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software
Brezina, V., T. McEnery, & S. Wattam	2015	#LancsBox [Computer Software], presented in 'A new perspective on collocation networks. Collocations in context'	20(2): 139–173	International Journal of Corpus Linguistics Available from http://corpora.lancs.ac.uk/lancsbox/
Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel	2014	The Sketch Engine: ten years on	1: 7–36	Lexicography Available from: http://www.sketchengine.eu
Scott, M.	2010	WordSmith Tools (v. 5).	v. 5	Oxford: Oxford University Press

Updated 2026-11-03