Subject Code	ENGL3006
Subject Title	Corpus-driven Language Learning
Credit Value	3
Level	3
Pre-requisite / Co-requisite/ Exclusion	None
Objectives	1. To understand basic concepts and issues in corpus linguistics. 2. To use corpora (large collections of electronic text) for a number of types of linguistic research. 3. To use corpus software programs and tools to discover patterns of language structure, use and meaning in different corpora. 4. To study how corpus linguistics has challenged conventional theories of language. 5. To study different applications of corpus linguistics in, e.g., dictionaries and grammars, register variation, language education and research, translation, critical language study, etc.

Intended Learning Outcomes	Upon completion of the subject, students will be able to:
	
	Category A: Professional/academic knowledge and skills
	
	a. apply corpus linguistic techniques to different types of corpora as databases to study a wide range of linguistic features;
	b. develop data-driven learning (DDL) strategies to establish a direct link between theories about language and the facts revealed by natural language, both individually and collaboratively
	c. report, in the form of an oral presentation and a written report, on a corpus-driven language study that students have conducted
	
	Category B: Attributes for all-roundedness
	
	d. develop analytical reasoning, critical thinking, and problem - solving skills:
	    • analytical reasoning: to think in a logical manner, supporting ideas with well-reasoned arguments and evidence
	    • critical thinking: to evaluate information and evidence critically, able to recognize flaws or inconsistencies in an argument

	• problem solving: to understand the problem, explore plausible answers, and select the most appropriate decision/solution for the problem e. employ a range of strategies for effective communication and learning autonomously and collaboratively; and f. develop intellectual curiosity to work both independently and as part of a team

Subject Synopsis/ Indicative Syllabus	1. Introduction to corpus linguistics
	2. Definition and use of a corpus
	3. Types of corpora
	4. Corpus software programs and tools, e.g. *WordSmith Tools Version 5* (Mike Scott), AntConC (Laurence Anthony), ConcGram (Chris Greaves), ParaConc, etc.
	5. Corpus search methods and analysis of corpus text: by word, phrase, collocate, concgram, keyness, and concordance
	6. The mechanics of corpus creation: Corpus design, construction and annotation
	7. Corpus-based and corpus-driven approaches to linguistic description and analysis
	8. A lexical approach to the description of English: Sinclair's (2004) five categories of co-selection (collocation, colligation, semantic preference, semantic prosody, and the core)
	9. Applications of corpora, e.g.
	  • the tracking of variation and change in the English language
	  • the production of dictionaries and other reference materials
	  • the study of all aspects of linguistics, including morphology, lexis, grammar, lexico-grammar, semantics, literal and metaphorical meanings, discourse structure, pragmatics, and discourse intonation
	  • the study of linguistic variation across modes (speaking and writing), registers (academic, business, social, scientific, legal, etc.) and genres (university textbooks, financial reports, conversation, laboratory reports, contracts, etc.)
	  • the study of multilingual and parallel texts

Teaching/ Learning Methodology	The teaching and learning approach is task-based, student-centred, interactive, and reflective.

Assessment Methods in Alignment with Intended Learning Outcomes	Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)					
			a	b	c	d	e	f
	1. Research critique	25%		✓	✓	✓	✓	✓
	2. Class activities	25%	✓	✓	✓	✓	✓	✓
	3. Project	50%	✓	✓	✓	✓	✓	✓
	Total	100 %						

All the assessed tasks allow the students to demonstrate how well they understand the subject knowledge, how effectively they can apply their knowledge of corpus processing and analysis, and how effectively they can conduct, report and critique corpus research.

Student Study Effort Expected	Class contact:	
	▪ Lectures	39 Hrs.
	▪ Seminars	0 Hrs.
	Other student study effort:	
	▪ Private study	58 Hrs.
	▪ Take-home assignments	29 Hrs.
	Total student study effort	126 Hrs.

| Reading List and References | **Recommended reading**

Biber, D., & Reppen, R. (Eds.). (2015). *The Cambridge Handbook of Engl Corpus Linguistics*. Cambridge: Cambridge University Press.
Crawford, W. J. (2016). *Doing corpus linguistics*. New York
Crosthwaite, P. (2020). *Data-driven learning for the next generation: Corp and DDL for pre-tertiary learners*. Abingdon, Oxon: Routledge.
Friginal, E. (2018). *Corpus linguistics for English teachers: new tools, onl resources, and classroom activities*. New York, NY: Routledge.
Ivor, T. (2015). *Corpus Linguistics for ELT: Research and Practice*: Taylo Francis. |

Lenko-Szymanska, A., Boulton, A., & ebrary, I. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning*. Amsterd John Benjamins.

London: New York: Routledge.

O'Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics*. London: Routledge.

O'Keeffe, Anne, McCarthy, Michael, and Carter, Ronald. (2007). *From Corpus* to Classroom. Language Use and Language Teaching. Cambridge: Cambridge University Press.

Pérez-Paredes, P., & Mark, G. (Eds.). (2021). *Beyond Concordance Lines: Corpora in language education*. Amsterdam: John Benjamins.

Reppen, R. (2010). *Using corpora in the language classroom*. New York.

Sinclair, John McH. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Sinclair, John McH. (2004). Trust the Text. London and New York: Routledge.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: a guide for resear* (First edition.. ed.). Abingdon, Oxon: Abingdon, Oxon: Routledge.

Further reading

Baker, P. (2006). Using Corpora in Discourse Analysis. London: Continuum.Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam; Philadelphia: John Benjamins.

Biber, D., Connor, U., Upton, T. A., Molly, A. and Gladkov, K. (2007). Rhetorical appeals in fundraising. In D. Biber, Ulla, C. and Upton, T.A. (Eds.), Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure (pp. 121-151). Amsterdam; Philadelphia: John Benjamins.

Cheng, W. (2006). Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. In J. Flowerdew and M. Mahlberg (Eds.), Special Issue of International Journal of Corpus Linguistics: Corpus Linguistics and Lexical Cohesion 11(3), 325-344.

Cheng, W., Warren, M. and Xu X.F. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. System 31(2): 173-186.

Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge: Cambridge University Press.

Revised by Phoebe Lin, August 2021