

Course Title	: Introduction to Corpus Linguistics
Course Code	: ENG3014
Recommended Study Year	: 3 rd or 4 th Year
No. of Credits/Term	3
Mode of Tuition	: Lecture-tutorial mode
Class Contact Hours	: 2-hour lecture per week 1-hour tutorial per week
Category	: Discipline Elective
Prerequisite(s)	: Nil
Co-requisite(s)	: Nil
Exclusion(s)	: Nil
Exemption Requirement(s)	: Nil

Brief Course Description

This course offers an introduction to corpus linguistics, a method that reveals patterns and insights about language by exploring real-world texts (corpora). Students will learn the essential concepts and techniques in corpus linguistics, including how to create, annotate, and analyse corpora. The course involves hands-on experience with online software tools for corpus research. Students will practice identifying linguistic patterns, examining language variation, and investigating discourse phenomena using corpus data. The course also highlights the applications of corpus linguistics in areas like language teaching, lexicography, and sociolinguistics. By combining theoretical insights with practical exercises, students will acquire skills to carry out independent corpus-based research projects.

Aims

- To introduce the fundamental concepts, principles, and methods of corpus linguistics.
- To familiarise students with key corpus resources and tools.
- To enhance students' skills in conducting corpus-based language studies.
- To expose students to current research in corpus linguistics and its applications in language teaching, lexicography, and sociolinguistics.

Learning Outcomes

Upon completing the course, students will be able to:

LO1	Identify and explain key concepts and fundamental methodologies in corpus linguistics, demonstrating a clear understanding of their applications.
LO2	Create, annotate and manage corpora tailored to specific research questions, showcasing proficiency in corpus design and data handling.
LO3	Utilise and evaluate tools and software for corpus linguistics research, assessing their effectiveness in various linguistic analyses.
LO4	Apply fundamental linguistic and statistical techniques for corpus analysis, interpreting results to draw meaningful conclusions.
LO5	Analyse and discuss corpus studies in the contexts of language teaching, lexicography and sociolinguistics, demonstrating the relevance of corpus linguistics in these fields.
LO6	Conduct independent research by formulating and testing hypotheses that are descriptively and theoretically sound, demonstrating critical thinking and analytical skills.

Indicative Contents

1. Introduction
 - What is corpus linguistics?
 - Historical development of corpus linguistics
 - Types of corpora
2. Corpus creation: Design, construction, and annotation
 - Principles of corpus design and sampling
 - Techniques for corpus construction
 - Annotation methods
3. Corpus description: Linguistic patterns
 - Methods for analysing linguistic patterns within corpora
 - Use of concordance tools to explore word usage and collocations
 - Frequency analysis and its implications for understanding language use
4. Corpus description: Language variation
 - Factors influencing language variation, such as region, social class, and age
 - Techniques for comparing language use across different corpora to identify variations
5. Corpus analyses
 - Analysis of specific research questions using corpora, such as language change over time
 - Interpretation of results and drawing conclusions from corpus data
6. The role of corpora in language teaching
 - How corpora inform language teaching methodologies and materials
 - Designing language teaching activities based on corpus data
 - Learner corpora analysis
7. The role of corpora in sociolinguistics
 - Using corpora to study language use in different social contexts
 - Analysing social variables and their effect on language through corpora

Teaching Methods

The course will include lectures, discussions, presentations and written assignments.

Measurement of Learning Outcomes (LOs)

Learning Outcome	Attendance and participation (Individual)	Corpus exercise (Individual)	Corpus creation (Individual)	Research project presentation (Individual)	Research project writeup (Individual)
LO1	X	X	X	X	X
LO2	X		X		
LO3	X	X		X	X
LO4	X	X		X	X
LO5	X			X	X
LO6	X		X	X	X

Assessment Methods	LOs					
	LO1	LO2	LO3	LO4	LO5	LO6
	√	√	√	√	√	√
	√		√	√		
	√	√				√
	√		√	√	√	√
	√		√	√	√	√

Assessment

Continuous assessment: 100%

- Attendance and participation (10%)
- Corpus exercise (15%):
This component involves hands-on practice with existing corpora. Students will use online concordancers to identify and analyse linguistic patterns within a given corpus. It is designed to familiarise students with the practical aspects of corpus analysis, ensuring they can navigate and interpret corpus data effectively.
- Corpus creation (25%):
Students will design, construct, and annotate their own small corpus. They will learn about the principles of corpus design, such as size, representativeness, and balance, and apply these in creating a corpus relevant to their chosen topic. This exercise is crucial for understanding the foundational steps in corpus linguistics research.
- Research project presentation (20%):
Students will present their findings from their corpus analysis in a formal presentation. This assesses their ability to communicate complex data and insights clearly and effectively. The presentation should include their research question, methodology, key findings, and implications, allowing for peer and instructor feedback.
- Research project writeup (30%):
The writeup is a detailed report of the research project, covering the entire process from corpus creation to analysis and interpretation. It should be structured like a research paper, with sections on introduction, methods, results, discussion, and conclusion. This component evaluates students' ability to write comprehensively about their research, demonstrating their understanding of corpus linguistics.

Required Readings

The handouts will supply most of the materials that students need. If they wish to consult further texts (e.g., for their own interest or for their presentation), the following resources may help.

Recommended/Supplementary Readings

- Cheng, W., & Lam, P. (2022). What can a corpus tell us about language teaching?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 299–312). New York, NY: Routledge, Taylor & Francis Group.
- Connor, U. & Upton, T. A. (Eds.). (2004). *Applied corpus linguistics: A multidimensional perspective*. Amsterdam: Rodopi.
- Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics*. New York, NY: Routledge.
- Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus*.

- Cambridge: Cambridge University Press.
- Lindquist, H., Levin, Magnus. (2018). *Corpus linguistics and the description of English* (2nd ed.). Edinburgh: Edinburgh University Press.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery T., & Wilson A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh: Edinburgh University Press.
- Meunier, F. (2021). Introduction to learner corpus research. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 23–36). New York, NY: Routledge.
- Meyer, C. (2002). *English corpus linguistics*. Cambridge: Cambridge University Press.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Sampson, G., & McCarthy, D. (Eds.). (2005). *Corpus linguistics: Readings in a widening discipline*. London, New York, NY: Continuum International Publishing Group.
- Sinclair, T. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Blackwell Publishers.
- Thomas, J. & Short, M. (Eds.). (1996). *Using corpora for language research*. London: Longman.
- Wynne, M. (2005). *Developing linguistic corpora*. Oxford: Oxbow Books.

Important Notes:

- (1) Students are expected to spend a total of 9 hours (i.e. 3 hours of class contact and 6 hours of personal study) per week to achieve the course learning outcomes.
- (2) Students shall be aware of the University regulations about dishonest practice in course work, tests and examinations, and the possible consequences as stipulated in the Regulations Governing University Examinations and Course Work. In particular, plagiarism, being a kind of dishonest practice, is “the presentation of another person’s work without proper acknowledgement of the source, including exact phrases, or summarised ideas, or even footnotes/citations, whether protected by copyright or not, as the student’s own work”. Students are required to strictly follow university regulations governing academic integrity and honesty.
- (3) Students are required to submit writing assignment(s) using Turnitin.
- (4) To enhance students’ understanding of plagiarism, a mini-course “Online Tutorial on Plagiarism Awareness” is available on <https://pla.ln.edu.hk/>.
- (5) Students must adhere to the University’s guidelines and practices when using Generative Artificial Intelligence (GAI) tools. The official documents “Guidelines for Using GAI Tools at Lingnan University” and “Best Practices for Ethical and Responsible Use of GAI Tools in Course Assessments” can be found here: <https://www.ln.edu.hk/tlc/generative-artificial-intelligence/gai-guidelines-best-practices>

Use of Generative Artificial Intelligence (AI) Tools in Course Assessments:

The University has allowed the use of ChatGPT and other Generative AI tools from the 2023-24 academic year. For this course, such use of AI shall be limited to information search only and must be declared in submitted assessments in accordance with the University’s regulation.