

The Hong Kong University of Science and Technology
UG Course Syllabus

Exploring Language and Communication with Big Data

LANG2090

3 credits

Pre-requisites/Co-requisites: N/A

Name of Course Coordinator: Dr. Yin Zhong

Email of Course Coordinator: lcyinzhong@ust.hk

Office Hours: Available by appointment.

Course Description

LANG 2090 is a three-credit, one-semester humanities (linguistics) broadening course for students of any discipline. Students attend two 1.5-hour classes per week over one semester. This course provides students with opportunities to use corpus tools to analyse language data, identify language patterns, and observe their applications in daily communication. Students will gain insights into language patterns through empirical studies, enhancing their digital literacy and equipping them with the quantitative skills necessary to interpret and leverage large datasets. Through the study of existing language corpora and the development of their own databases, students will deepen their understanding of how language impacts communication and societal dynamics.

Course Outline

Module		Lesson	Course Topics
1	Foundations of Corpus Linguistics	1	What Is Corpus Linguistics and Why Does It Matter?
		2	Finding Patterns: Keywords and Frequency
		3	Meaning in Context: Collocations, Semantic Preference, and Semantic Prosody
		4	From Lines to Insights: Comparing Corpora Across Regions, Genres, and Time
		5	From Curiosity to Corpus: Proposing a Topic for Inquiry Assessment 1 Consultation
2	Applying Corpus Methods to Real-World Communication	6	Domain Deep Dive I: Health and Science Communication
		7	Submission of Assessment 1
		7	Domain Deep Dive II: Business and Political Communication
		8	Domain Deep Dive III: Digital and Media Communication
		9	From Interest to Inquiry: Developing a Corpus-Based Research Question
		10	From Topic to Texts: Building Your Dataset
		11	From Raw to Ready: Cleaning and Annotating Your Corpus
		12	Building a Story with Data: Structuring and Writing Your Analysis
		13-14	Assessment 2 Consultation Submission of Assessment 2

Methods of instruction

Interactive tutorials with group discussions, hands-on data analysis seminars, peer collaboration, instructor-led input, and guided projects.

Intended Learning Outcomes (ILOs)

By the end of the course, students should be able to:

1. Explain how corpus linguistics and language data can provide insights into real-world communication.
2. Formulate focused and feasible research questions about language use in specific communication domains.
3. Use basic corpus methods and digital tools to analyse textual datasets.
4. Identify, interpret and visualise meaningful patterns in linguistic data.
5. Communicate evidence-based findings clearly in writing, supported by accurate and effective data visualisations.
6. Work collaboratively to plan and carry out a corpus-based language analysis project.
7. Evaluate the social and ethical implications of collecting and using language data.

Assessment and Grading

	Assessment Task	Contribution (%)	Due Date
1	Individual Analytical Essay <i>Students will write a 1,200-word analytical essay comparing patterns of word use and meaning across corpora using appropriate corpus evidence.</i>	35%	Week 7*
2	Group Corpus Project <i>In groups of 3–4, students will write a 3500-word report analysing linguistic trends within a communication domain of their choice using corpus-based methods.</i>	45%	Week 14*
	Peer Evaluation <i>Group members will assess each other's contributions after the completion of the group project.</i>	5%	Week 14*
3	Course Participation <i>Assessed through attendance, engagement in class discussions, and completion of hands-on in-class activities.</i>	15%	Throughout the semester

*Specific due dates are posted on Canvas.

Mapping of Course ILOs to Assessment Tasks

Assessment Task	Mapped ILOs	Explanation
Individual Analytical Essay	ILO1, ILO2, ILO3, ILO4, ILO5	Assesses students' ability to formulate a focused question, apply corpus methods to compare language patterns across contexts, interpret relevant evidence, and communicate their findings clearly in writing.
Group Corpus Project	LO1, ILO2, ILO3, ILO4, ILO5, ILO6, ILO7	Assesses students' ability to plan and conduct a collaborative and domain-specific corpus analysis; compile and process textual data; interpret and visualise findings; communicate the results in a coherent written report; and consider

		relevant social, ethical and methodological issues.
Peer Evaluation	ILO6	Assesses each student's contribution to and collaboration within the group project.
Course Participation	ILO1, ILO3, ILO4, ILO6, ILO7	Measures students' ongoing application of corpus concepts, methods and digital tools through hands-on, collaborative and ethically informed in-class activities.

Grading Rubrics

Detailed rubrics for each assignment are provided on Canvas.

Final Grade Descriptors

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Knowledge, analysis and digital literacy: Demonstrates outstanding understanding of corpus concepts and methods. Selects and processes data carefully, uses corpus tools accurately, and produces insightful and critically evaluated interpretations. Recognises relevant methodological limitations and ethical considerations. Communication of evidence: Findings are clearly and coherently structured in writing and strongly supported by appropriate corpus evidence. Tables, figures and other data visualisations are accurate, effective and well-integrated into the analysis.
B	Good Performance	Knowledge, analysis and digital literacy: Demonstrates a solid understanding of corpus concepts and methods. Uses appropriate tools and generally interprets the results accurately. Some critical evaluation and synthesis are evident, although parts of the analysis could be developed further. Communication of evidence: Findings are logically organised in writing and supported by relevant corpus evidence. Visualisations are generally clear and appropriate but may occasionally lack precision or effective integration.
C	Satisfactory Performance	Knowledge, analysis and digital literacy: Demonstrates a basic understanding of corpus concepts and methods. Uses corpus tools appropriately in most cases, but the analysis remains mainly descriptive, with some misinterpretation or limited critical insight. Communication of evidence: Findings are communicated adequately in writing but may not always be clearly organised or sufficiently supported. Visualisations are included but may be inconsistent, underdeveloped or only partly connected to the analysis.
D	Marginal Pass	Knowledge, analysis and digital literacy: Demonstrates limited understanding of corpus concepts and methods. Application of tools is inconsistent, and claims may be inaccurate or poorly supported. Analysis and critical interpretation are minimal.

		Communication of evidence: Written findings lack clarity or logical organisation. Corpus evidence and visualisations are limited, difficult to interpret or poorly integrated into the analysis.
F	Fail	Knowledge, analysis and digital literacy: Demonstrates serious misunderstanding or misuse of corpus concepts, data or tools. The work lacks meaningful analysis, critical interpretation or sufficient evidence and does not meet the main task requirements. Communication of evidence: Written findings are incoherent, unsupported or absent. Corpus evidence and visualisations are missing, inaccurate or misleading.

Course AI Policy

Use of GenAI tools is permitted for idea generation, language suggestions, and brainstorming; however, students must critically evaluate outputs and ensure originality in assessed work. Proper citation of AI tools is required when used.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within ten working days of submission. Feedback on assignments will include strengths and areas for improvement where relevant. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

Resubmissions are not accepted unless under exceptional circumstances.

Required Texts and Materials

Course materials and additional resources are provided via Canvas.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity | HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

Plagiarism

In CLE assignments, students should not copy from others nor submit the same work for multiple courses. Source materials must be explicitly acknowledged.

Assignments with evidence of plagiarism may score a zero and be followed up with your department's Dean. To demonstrate academic integrity and avoid plagiarism, you are expected to:

- paraphrase, summarize, reference and synthesize ideas from sources
- refrain from copying expressions and ideas directly from any source (e.g. from another student, or from a video, or from an article, etc.) without proper acknowledgement