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 Leader: Dr. Yin Zhong
Email of Course Leader: 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 Registers, 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 Scientific Communication
			Submission of Assessment 1
		7	Domain Deep Dive II: Business and Political Communication
		8	Domain Deep Dive III: Digital and Media Communication
		9	From Topic to Texts: Building Your Dataset
		10	From Raw to Ready: Cleaning and Annotating Your Corpus
		11	Building a Story with Data: Structuring Your Analysis
		12	Assessment 2 Consultation + Presentations of Assessment 2
		13	Presentations of Assessment 2 & Wrap-up
			Submission of Assessment 2

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

Intended Learning Outcomes (ILOs)
By the end of the course, students should be able to:
1. Understand and explain how corpus linguistics and language data can inform real-world communication.
2. Formulate research questions related to language use in specific domains.
3. Analyze large-scale textual data using basic corpus methods and tools.
4. Interpret linguistic data using visualization and summary techniques.
5. Communicate findings clearly through writing, oral presentation, and visual means.
6. Work collaboratively to plan and carry out language data analysis projects.
7. Reflect on the social and ethical implications of using big data in language research.

Assessment and Grading

	Assessment Task	Contribution (%)	Due Date
1	**Individual Analytical Essay** *Students will write a 1200-word analytical essay comparing word usage and meanings across corpora.*	30%	Week 6*
2	**Group Project Report** *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.*	35%	Week 13*
	Group Project Presentation *Each group will deliver a 10-min presentation summarizing their project report, followed by a 5-min Q&A session.*	15%	Week 13*
3	**Course Participation** *Assessed through attendance, engagement in group discussions, and completion of hands-on in-class activities.*	15%	Throughout the semester
4	**Peer Evaluation** *Group members will assess each other's contributions after the completion of the group project.*	5%	Week 13*

*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, ILO5	Demonstrates the ability to analyse word usage patterns across contexts and communicate findings in written form.
Group Project Report	ILO1, ILO2, ILO3, ILO4, ILO5	Assesses students' skills in conducting domain-specific corpus analysis and reporting findings.
Group Project Presentation	ILO1, ILO2, ILO3, ILO4, ILO5	Assesses presentation and collaboration skills, along with the ability to communicate data insights clearly.
Course Participation	ILO1, ILO2, ILO3, ILO4, ILO5	Measures ongoing engagement in discussions and hands-on tasks.

Peer Evaluation	ILO2, ILO3, ILO4, ILO5	Measures teamwork, contribution, and collaboration.

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	Content: Demonstrates outstanding understanding and depth of analysis. Corpus data is carefully selected and interpreted. Shows strong critical evaluation of linguistic patterns and corpus tools. Insightful synthesis across contexts or corpora. Communication: Argument and narrative are clear and compelling. Strong integration of corpus evidence. Visuals and structure enhance audience understanding. In presentations, delivery is confident, fluent, and well-paced.
B	Good Performance	Content: Demonstrates a solid grasp of corpus methods and key linguistic features. Mostly appropriate use of tools and interpretation of results. Some critical evaluation is present. Synthesis is logical but not always fully developed. Communication: Mostly clear and logical argumentation. Data is presented in readable form but may lack impact or precision. Writing/speaking is generally effective. Minor lapses in structure, transitions, or delivery.
C	Satisfactory Performance	Content: Meets basic expectations. Corpus data is used appropriately but analysis remains descriptive. Some misinterpretation or surface-level synthesis. Critical insights are limited. Communication: Communication is adequate but lacks polish. Arguments are occasionally hard to follow. Visuals or structure may be inconsistent or underutilized. Oral or written delivery is uneven.
D	Marginal Pass	Content: Limited understanding of corpus methods. Inaccurate or poorly supported claims. Minimal critical thinking. Synthesis is weak or missing. Communication: Communication lacks clarity or logical progression. Data is hard to interpret or poorly integrated. Visual aids or structure may confuse rather than clarify. Weak or unclear oral delivery.
F	Fail	Content: Misuse or misunderstanding of corpus data. Lacks analytical depth and fails to meet task requirements. No critical evaluation or synthesis. Communication: Communication fails to convey key ideas. Argument is incoherent or unsupported. Visuals are missing or misleading. Oral delivery is ineffective or absent.

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