

**City University of Hong Kong**

**Information on a Course  
offered by Department of English  
with effect from Semester A in 2012/2013**

---

**Part I**

Course Title:	Corpus Approaches to Language Studies
Course Code:	EN6506
Course Duration:	1 semester
Credit Units:	3
Level:	P6
Medium of Instruction:	English
Prerequisites:	Nil
Precursors:	Nil
Equivalent Courses:	Nil
Exclusive Courses:	Nil

**Part II**

**Course Aims:**

This course aims to provide participants with an understanding of the theories, concepts and skills needed to develop, analyse, and use electronic text collections (corpora) of English. It thus offers both theoretical and practical perspectives on the topic, beginning with a brief introduction to the history of language corpora, then quickly moving on to identifying some of the basic practical skills needed to create and use corpora, all through hands-on computer practice. Participants will discover corpus-linguistic approaches to the collection, organisation and analysis of authentic written and spoken texts drawn from a variety of genres and purposes. Throughout the course, participants will be able to apply their knowledge and practical skills to critical reflection upon theoretical and pedagogical issues, principles, and techniques related to the use of language corpora for teaching and research. Topics will also include introductions to the organisational principles of language resources, search tools and software for corpus research, the applicability of corpus data to pedagogical purposes, and the utility of corpora for language analysis and description, e.g. as an ideal basis for writing research papers or the dissertation.

Pre-requisites: basic computing abilities and strong interest in learning how to use (user-friendly) computer programs for linguistic analysis. Additional skills will be taught on the course at a relatively gentle pace.

## Course Intended Learning Outcomes (CILOs)

*Upon successful completion of this course, students should be able to:*

No.	CILOs
1.	Identify a number of key corpus-linguistics concepts and some available language corpora (along with their basic design, organisational principles, purpose, contents, uses, type of language variety they represent, differences among them, their strengths and weaknesses).
2.	Create self-compiled special-purpose text collections (corpora) for language learning, teaching and research, according to the design principles and corpus-theoretic methods taught.
3.	Analyse the linguistic features of various general and specialised genres of English following corpus-linguistic theories, methods and techniques.
4.	Reflect on issues and problems in the application or use of corpus-linguistic theories and methods by researchers, teachers, and learners.
5.	Apply corpus-linguistic theories and methods to the linguistic analysis of self-compiled & currently available corpora of authentic written and spoken texts, using suitable software, web tools and systematic analytical techniques.

## Teaching and Learning Activities (TLAs)

*(Indicative of likely activities and tasks designed to facilitate students' achievement of the CILOs. Final details will be provided to students in their first week of attendance in this course)*

This section is indicative of likely activities and tasks students will undertake to learn in this course. Final details will be provided to students in their first week of attendance in this course.

CILO No	TLAs	Hours/week
CILOs 1 – 5	<p><b>Interactive lectures, using web-based resources and awareness-raising in-class exercises</b></p> <p>All course activities centre around the web-based teaching materials that will be discussed step-by-step in an interactive lecture, accompanied by various activities that provide participants with the opportunity to interact with the instructor and each other in group discussions and in-class exercises using different types of software in order to:</p> <ul style="list-style-type: none"> <li>(i) discover key corpus-linguistics concepts and existing language corpora (along with their basic design, organisational principles, purpose, contents, uses, type of language variety they represent, differences among them, their strengths and weaknesses, etc.),</li> <li>(ii) analyse linguistic features of general and specialized genres of English using corpus-linguistic methods,</li> <li>(iii) gain sufficient knowledge and practical experience to reflect on issues and problems in the application or use of corpus-linguistic theories and methods,</li> <li>(iv) learn how to create self-compiled corpora for their own analyses.</li> </ul>	1-3 hours/week, over 12 weeks

## Assessment Tasks/Activities

(Indicative of likely activities and tasks designed to assess how well the students achieve the CILOs. Final details will be provided to students in their first week of attendance in this course)

This section is indicative of likely activities and tasks students will undertake to learn in this course. Final details will be provided to students in their first week of attendance in this course.

CILO No	Type of assessment tasks/activities	Weighting	Remarks
CILOs 1 & 2	<b>Assignment 1 – Creating basic corpora</b> In this assignment, students should demonstrate a basic understanding of how corpus materials can be collected and compiled from text archives and other web sources. They should furthermore illustrate some initial steps they have identified and taken in organising and partially ‘cleaning up’ this data in order to make it suitable for analysis.	30%	Individual work; assessed individually
CILOs 3 – 5	<b>Assignment 2 – Discovering &amp; analysing/interpreting language patterns</b> This assignment has two separate main parts, one where students demonstrate their understanding of identifying and analysing language patterns by making use of their knowledge of regular expressions they will have acquired on the course, and the other where they perform an analysis of the lexical features of a number of texts in order to gain insights into textual complexity or collocations.	35%	Individual work; assessed individually
CILO 2 – 5	<b>Assignment 3 – Creating &amp; exploring linguistically annotated data</b> For this assignment, students get to select a small number of files from a learner corpus and, based on what they have learnt in class, devise their own annotation scheme to mark up features relevant interesting features of learner interlanguage they discover through basic exploration. Once they have created the basic annotation, they then explore the data further by concordancing on and evaluating/revising their own annotations, as well as conducting other types of analysis on the chosen texts.	35%	Individual work; assessed individually

### Grading of Student Achievement:

Refer to Grading of Courses in the Academic Regulations for Taught Postgraduate Degrees.

Grading pattern: Standard (A+, A, A-....F). Grading is based on student performance in assessment tasks/activities.

## Part III

### Keyword Syllabus:

- language corpora (spoken and written)
- corpus design, compilation and annotation (encoding formats; headers; XML mark-up; tags)
- corpus analysis tools (concordancers; taggers)
- word (frequency) lists
- collocations; key words
- corpus-based materials design and pedagogy

### Additional Reference Textbooks:

Hoffmann, Sebastian, Evert, Stefan, Smith, Nicholas, Lee, David YW and Ylva Berglund Prytz. (2008). *Corpus linguistics with BNCweb—A practical guide*. Frankfurt am Main: Peter Lang. [PE1074.5 .H64 2008]

Kennedy, Graeme D. (1998). *An introduction to corpus linguistics*. London: Longman. [P98 .K44 1998]

Meyer, Charles. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press. [PE1074.5 .M49 2002; + [online access from ebrary](#)]

Other resources & materials are referenced from within the online course materials.

### Online Resources

Course materials available online (from CityU computers) at [http://www9.english.cityu.edu.hk/martin\\_weisser/corpus\\_approaches/corpus\\_approaches.html](http://www9.english.cityu.edu.hk/martin_weisser/corpus_approaches/corpus_approaches.html)