

Special Topics in Functional English

"Introducing Corpus Methods in Data Science, Digital Humanities, and Applied Linguistics"

Instructor: Laurence Anthony

Course Outline

In this course, students will learn how to create and analyze databases of language (corpora) and visualize and interpret the results of that analysis for use in data science, digital humanities, and applied linguistics, applications. The course will be divided into three parts. In part one of the course, students will learn how to access publicly available language databases and create their own custom language datasets through specialized search tools, web-page scraping, and file conversion programs. In part two of the course, the language datasets introduced in part one will be used to introduce basic corpus methods for finding high frequency word and phrase patterns, locating where and when language patterns appear in written and spoken texts, visualizing interactions between different speakers or writers, and interpreting the analyses and visualizations of language. In part three of the course, students will apply what they have learned in parts one and two of the course in an analysis, visualization, and interpretation of language in their own area of interest, for example, scientific writing, social network communication, news media, or the language of anime, film, and television.

Objectives

1. Learn how to access publicly available language data (corpus) resources.
2. Learn how to build custom language database (corpus) resources.
3. Learn basic corpus methods to analyze language and visualize the results of that analysis.
4. Learn how to interpret analyses and visualizations of language.
5. Apply corpus methods in the analysis, visualization, and interpretation of language in a particular area of student interest.

Assigned work before/after class

Before the course, students should be familiar with the basic design of a research project as explained in the Concept Building and Discussion (CBD) course. On completing the course, students should be ready to take a course in technical writing or presentation and incorporate the data analyses methods from this course in their research.

Course Schedule

Week 1	General Introduction: Aims of course. Evaluation procedure. Impact of corpus methods on data science, digital humanities, and applied linguistics.
Lesson 2	Accessing and using publicly available language resources
Lesson 3	Creating custom language resources through specialized search tools, web-page scraping, and file conversion programs
Lesson 4	Using custom language resources to understand how language is used in the world
Lesson 5	Introduction to primary corpus methods in the analysis of language
Lesson 6	Applications of corpus methods in applied linguistics - Part 1: (vocabulary and grammar)

Lesson 7	Applications of corpus methods in applied linguistics - Part 2: (stylistics and pragmatics)
Lesson 8	Applications of corpus methods in digital humanities - Part 1: (news, politics, and law)
Lesson 9	Applications of corpus methods in digital humanities - Part 2: (film, television, and popular culture)
Lesson 10	Applications of corpus methods in data science - Part 1: (collecting and cleaning big data)
Lesson 11	Applications of corpus methods in data science - Part 2: (analyzing and visualizing big data)
Lesson 12	Student project - Part 1: Research question formulation and background reading
Lesson 13	Student project - Part 2: Data collection and analysis
Lesson 14	Student project - Part 3: Analysis and visualization of results
Lesson 15	Review lesson and final paper check

Textbook

Specially prepared materials will be distributed in each class of the course.

Reference Materials

None

Grading

Students will be evaluated based on a portfolio of work comprising of:

1. Final paper (>1500 words) that demonstrates an understanding of corpus methods as applied in the analysis of language in an areas of interest of the student. (60%)
2. In class review quizzes and homework exercises (30%)
3. In-class participation (10%)

Notes

1. The lectures will be conducted in English and all materials will be distributed in English.
2. Students are expected to understand English lectures and participate in classroom discussions in English.
3. No prior knowledge of corpus methods, data science, digital humanities, or applied linguistics is expected.