

English 4508: Corpus Linguistics and Language Analysis

Corpus Linguists conduct computerized or computer-assisted language research using corpora, which are large, carefully constructed collections of naturally occurring language events (texts or transcripts). This entails both qualitative and quantitative methods to address questions of language use patterns, with implications for linguistic theory, second language education, and a considerable number of interdisciplinary and real-world applications in Business, Law, Economics, Political Science, and more. These methods of inquiry are even more valuable today with GenAI resources putting language technologies in the forefront of the public eye.

This course emphasizes the methodological considerations involved in corpus construction, annotation, and analysis within a broader picture of linguistic theory. There is careful attention paid to real-world applications. Students will develop an understanding of the motivations and affordances of a corpus-based approach to language analysis, familiarity with freely available corpora (e.g., the Corpus of Contemporary American English) and related tools, basic skills in corpus compilation and annotation (e.g., part-of-speech tagging, syntactic parsing), and an understanding of corpus-based analysis. No background in corpus-linguistics is required, and considerable resources will be introduced for future study (for those interested) in python, NLP, and statistics.