

Tutorial: Using corpora at GU (February 10, 12:30-13:45, Poulton Hall 230)

Corpora are searchable collections of digital texts, often containing a large amount of language data for research in linguistics and other disciplines. To help make corpus resources at Georgetown more accessible, this tutorial focuses on the basics of using the GU CQP web interface, which indexes over 4.8 billion word forms in more than 40 corpora at present, almost all of which contain automatic part-of-speech analysis and lemmatization, as well as some metadata about each text. The materials come from a variety of sources and genres, including newspaper data, literature, language from the Internet, transcribed conversation, essays written by L2 learners and native speakers, political speeches and more. No previous technical knowledge is required to participate, but please send an e-mail to amir.zeldes@georgetown.edu if you would like to register