**Introduction to Methods in Corpus Linguistics**

**Professor:** Kristopher Kyle

**Contact Information:** [kristopherkyle1@gmail.com](mailto:kristopherkyle1@gmail.com)

**TA:** Ji-hyun Jeon, [wjswlgus6623@gmail.com](mailto:wjswlgus6623@gmail.com)

**Location:** Oesol Hall 326-1

**Meeting Time:** 9:30am-12:30pm

**Course Duration:** August 1-August 14th, 2019

**Course overview**

This course provides a hands-on introduction to conducting research using corpora. Students will gain functional knowledge of foundational corpus analysis techniques using freely available software (e.g., AntConc and Python).

Each class will begin with a discussion of the theoretical and practical issues related to the topic of the day. Students will then be guided through a corpus analysis method in a structured format. After students have mastered the use of a particular method, they will then apply that method to a new situation (according to each student’s research interests). Accordingly, this course will cater to students whose research focuses on issues in linguistics, applied linguistics and/or literary studies.

**Textbook**

The readings for this course include research articles and book chapters curated by the professor. All readings will be made available online.

**Other materials**

Students must bring a laptop to each class. This is explicitly a hands-on course, and much of our time in class will be spent conducting computer-based analyses.

**Assignments**

In this course, there will be four projects in which you will demonstrate your ability to apply the skills learned in class to new corpora/problems. Note that projects can be completed in small groups (of no more than three people). The deadlines for each are outlined below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Project Name | Skills/Topics | Deadline | Expected Length | Percent of Grade |
| Mini-Project 1 | Corpus design, KWIC analysis, keyword Analysis | 2019-08-05 | 1500-2000 words | 25% |
| Mini Project 2 | Collocation and *n*-gram analysis | 2019-8-09 | 1250-1500 words | 20% |
| Mini Project 3 | Corpus tagging and analysis | 2019-8-12 | 1250-1500 words | 20% |
| Mini Project 4  Outline | Rough project plan | 2019-8-14 | 300-500 words | 5% |
| Mini Project 4 | Application Project | 2019-8-30 | 3000-4000 words | 30% |

**Course Schedule**

Please see the table below for the tentative course schedule

|  |  |  |
| --- | --- | --- |
| **Session** | **Topic(s)** | **Assignments** |
| W1.1 | Introduction to corpus design and analysis and AntConc Primer (Frequency; KWIC) | **Before Class:**   * Install AntConc * Read McEnery & Hardie (2011) Chp 1   **In Class:**   * Conduct frequency analysis * Conduct 3 KWIC analyses |
| W1.2 | Investigating similarities and differences between corpora: Keyness analysis | **Before Class:**   * Gabrielatos (2018)   **In Class:**   * Extract keywords from various corpora using AntConc * Conduct follow up analyses |
| W2.1 | Investigating fixed patterns: Bundles, clusters and *n-grams* | **Before Class:**   * Read Biber, Conrad, & Cortes (2004) * Read Mahlberg (2013) Chp 3   **In Class:**   * Extract n-grams from various corpora using AntConc * Conduct follow up analyses |
| W2.2 | Investigating related words: Collocation analysis  ­ | **Before Class:**   * Read Hunston (2002) Chp 4   **In Class:**   * Conduct collocation analyses using AntConc * Conduct follow up analyses |
| W2.3 | Introduction to corpus analysis with Python | **Before Class:**   * Install [**Anaconda Version of Python 3**](https://www.anaconda.com/download/)   **In Class:**   * Python primer * Replicate word and n-gram frequency analyses using Python |
| W2.4 | Dealing with messy texts: Cleaning and manipulating corpora | **Before Class:**   * Install Spacy package for Python * Read Kyle (2020, pp. 454-457)   **In Class:**   * Clean a messy corpus * Lemmatize and familize a corpus |
| W2.5 | Annotate corpora for part of speech | **In Class:**   * Annotate a corpus with fine-grained part of speech tags * Annotate a corpus with universal part of speech tags * Conduct part of speech specific corpus analyses (frequency, etc) |
| W3.1 | Annotate corpora for syntactic relationships | **In Class:**   * Annotate a corpus with syntactic dependencies * Annotate a corpus with universal part of speech tags   Conduct part of speech specific corpus analyses (frequency, etc) |
| W3.2 | Topics: TBD by class interests  [Creating lists for pedagogical purposes] | **Possible Readings:**   * Nation (2016) Chp 14 * Dang, Coxhead, & Webb (2017) |
| W3.3 | Topics: TBD by class interests  [Literary Stylistics] | **Possible Readings:**   * Mahlberg (2013) Chp 1, 2 |