Corpus Linguistics and English Education

(Graduate Course, Spring 2024)

- Course Schedule: 7~10PM (Wednesdays)
- Instructor: Miran Kim (mirankim316@gmail.com)
- Online classroom: https://MK316.github.io

❦ Course overview

This graduate-level course, designed for in-service English teachers at the secondary education level, offers an insightful exploration into corpus linguistics, combined with a practical introduction to Natural Language Processing (NLP) using basic Python coding. By integrating these computational techniques, the course aims to enhance the study of large electronic text collections (corpora) and their application in understanding language use and patterns. Participants will gain a deeper comprehension of language variation and how it can inform and improve teaching practices, leading to more effective development of teaching materials and activities. The course provides a balanced mix of theoretical instruction and practical application, focusing on the analysis and interpretation of corpus data in English language usage and introducing essential NLP techniques through Python programming.

❦ Learning Objectives

Upon completion of this course, participants will be able to:

1. Understand the principles of corpus linguistics and basic NLP.
2. Acquire foundational Python coding skills for text analysis.
3. Be able to apply corpus and NLP methods to develop effective English language teaching strategies and materials.
4. Enhance their capability to teach aspects like vocabulary, grammar, and comprehension through data-driven insights.

❦ Required Materials

- Textbook: [1] Selected chapters from McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*, Cambridge Textbooks in Linguistics. [2] Jump to Python [3] Supplementary readings (online)
- Python coding (Github, Colab)

❦ Course Requirements & Evaluation

1. Textbook reading and in-class discussions (30%)
2. Mini-projects and presentations (30%)
3. Language teaching app design (30%)
4. Course Participation (10%)

❦Course Policy

1. **Attendance Policy**: Regular attendance is essential for successful completion of this course. Each absence will result in a deduction of 5 points from the student's overall grade. Students are encouraged to attend every session to fully engage with the course material and participate in class discussions.
2. **Use of Generative Artificial Intelligence in Classwork and Assignments**: Students are encouraged to actively utilize generative AI technologies in their coursework and assignments. However, it is mandatory to clearly cite the source and extent of AI usage. This transparency is crucial for maintaining academic integrity and ensuring the responsible use of AI tools. The use of AI should complement, not replace, the student's own analysis and critical thinking.

Key Topics:

- **Introduction to Corpus Linguistics**: Exploring the design and compilation of text corpora.
- **Basic Python Coding for NLP**: Introduction to Python programming for text analysis. Simple coding exercises to familiarize with Python syntax and libraries like NLTK.
- **Word Frequency Analysis**: Using Python to automate the process of identifying and analyzing the frequency of words in a text corpus.
- **Collocation and Concordancing**: Learning to use Python scripts to discover common collocations and conduct concordance analysis in large text corpora.
- **Corpora in Classroom Teaching**: Applying findings from corpus linguistics and NLP to enhance teaching in vocabulary development, grammar, writing, and reading comprehension.
- **Hands-on Practice (Language Application Design)**: Participants will engage in practical exercises and mini-projects, using Python to manipulate and analyze corpus data relevant to their teaching contexts.

Weekly Schedule and assignments

There will be reading assignments (TBA)

Weeks	Topic(s)	Assignments
Week 01	Course overview	
Week 02	Introduction to corpus linguistics; Online Corpora	
Week 03	Data types, nltk	
Week 04	Nltk, word cloud, word frequency list	
Week 05	Type vs. token, lemmatization, nltk	
Week 06	Text analysis, words in context, concordance, collocations	
Week 07	TTR and other lexical diversity measures	
Week 08	Mini-project	
Week 09	Readability measures	
Week 10	NLP preprocessing, topic-modelling	
Week 11	Data collection, Sentiment Analysis	
Week 12	Data collection, Clustering Analysis	
Week 13	Project ideas, individual project discussions, samples	
Week 14	Individual project discussions, samples	
Week 15	Language Learning Applications (Final presentation)	

* Week 11: May 15 (W) Holiday