Assignment Guide for Mini-Project 3
Introduction to Methods in Corpus Linguistics
(updated 2021-07-22)

Purpose of project

To demonstrate functional knowledge of common analyses in corpus linguistics using Python,
including:
   a) Tag a corpus
   b) Conducting a frequency analysis
   c) Conduct a keyness OR collocation analysis with the tagged corpus

**NOTE: All files for Mini-Project 3 should be placed in a single folder and compressed before
submission. Please submit a .zip file with "MP3_Your_Name.zip" (e.g., "MP3_KrisKyle.zip" as
the filename.**

Procedure:
Step 1: Corpus description

You can use any corpus (other than the Brown corpus). You will report the following
information about the corpus you are using

Be sure to report:
   a) The language use domain that your corpus represents
   b) The number of words in your corpus
   c) The number of documents included in your corpus
   d) How you collected the data
   e) Benefits and limitations of your corpus

Step 2: Tag your corpus!

Using the Python functions we have worked on in class (see the corpus-toolkit page:
https://kristopherkyle.github.io/corpus_toolkit/) tag your corpus using Penn, Universal POS or
Dependency tags. Be sure to indicate which you chose (and why).

Write your tagged corpus to a new directory, and include it in your assignment submission.

Step 3: Conduct a frequency and range analysis using Python

Conduct a frequency analysis on your tagged corpus in Python and report the top 20 most
frequent items. Also conduct a range analysis and report the top 20 items

Step 4a: Conduct a keyword or collocation analysis in Python using your tagged corpus

Conduct a keyness analysis between your tagged corpus and a reference corpus (you can used a tagged version of the Brown corpus as your reference) OR conduct a collocation analysis with a content word that is relatively frequent in your corpus.

Report your top 20 key tagged words or your top 20 tagged collocates

Step 4b: Conduct a follow-up concordance analysis on one of the tagged items from your keyness or collocation analysis.

Using the tagged corpus that you wrote to a new directory in Step 1, conduct a follow up concordance two of your keywords/collocates. You can use Python (see Python Tutorial 5: https://kristopherkyle.github.io/corpus-analysis-python/Python_Tutorial_5.html or use AntConc).

Step 5: Submit your MP3 write up, Python script (i.e., the code you ran to complete the steps in this Mini-Project), original corpus, and tagged corpus in a .zip file to me via email by Thursday, July 29th before class begins at 10:00am.

Mini-Project 3

Introduction
<insert introduction here – provide an overview of the project and indicate why the analysis you are conducting might be interesting/useful>

Corpus Description
<Describe characteristics of your corpus here. Be sure to outline the target language use domain and the purpose of the corpus.>

Tagged Corpus Choices
<Include a brief description of your tagged corpus, including the choices you made (e.g., lemmatized or not, lowered or not, type of tag used).>

Report Most Frequent Tagged Words
<Include a table here>

Report Words with Highest Range Values
<Include a table here>

Brief Description of Keyness or Collocation Analysis
<Include a brief description of your analysis. In addition, include a table here>

Keyword or Collocate Use Analysis
<Report the findings of your use analysis here.>

Conclusion
<Summarize your results here. Also, explain how much you love Python ;) >