Text Analysis Using Lexos

[Click to schedule a meeting with the DITI Team]

Important Terminology

- **Corpus** (plural: corpora): a text or collection of texts that is used for analysis.
    - Example: One could create a corpus of all of Frederick Douglass's speeches to trace his language use over time.
- **nGram**: A continuous sequence of *n* items in a text.
    - For example, in Frederick Douglass's speeches, a bigram (or 2 continuous tokens) could be 'United States' and a trigram (3 tokens) could be 'justice for all'.
- **Stop words**: words that appear frequently in a language, like pronouns, prepositions, and basic verbs. These are often removed for computational analysis.
    - Some English stopwords include: a, the, she, he, I, me, us, of, is, would, could, should, etc.
- **Computational Text Analysis**: Text analysis is making inferences based on textual data. Computational text analysis (CTA) involves a computer drawing out patterns in a text, and a researcher interpreting those patterns. CTA includes methods such as word count frequency, nGrams, and sentiment analysis. CTA is similar to statistical analysis, but the data are texts.
- **Word Count Frequency:** Counting the total times a word appears in a text/corpus or the percentage of how often it appears

[Click here to navigate to the Lexos website]

What is Lexos?

Lexos is a web-based tool to help explore different corpora of digitized texts. Lexos provides a workflow of practices to help you be cognizant of choices you are making. For Lexos, separate .txt files are required. Other supported file formats include .txt, .html, .xml, .sgml, and .lexos.

Lexos provides a step-by-step guide for corpus uploading, preparation, and analysis.

Developed by: Colleen Nugent and Sayyara Poliferno
Questions? Contact us: nulab.info@gmail.com

Using Lexos:

1. Upload
    a. Upload your corpus
    b. Click “Browse” and select your entire corpus
2. Manage
    a. Make sure all documents in the corpus you want to use are selected.
        i. Blue = selected
        ii. Gray = not selected
3. Prepare
    a. Click “Prepare” and then “Scrub,” which transforms the text in your corpus and makes choices that will impact your results. Some possibilities:
        i. Make all your letters lowercase (the computer reads upper and lower case letters as two separate characters).
        ii. Remove punctuation, which may influence results.
        iii. Remove a list of words, usually stopwords.
    b. Once you have finished your text preparations, click “Apply” and wait a few minutes.
    c. You can also download your corpus with the transformations you chose applied for use with other text analysis tools.
4. Visualize
    a. Click “Visualize” and a drop down of visualization choices will appear.
    b. Click “Word Cloud” to visualize a word cloud across the entire corpus.
    c. Click “Multi Cloud” to visualize word clouds for each individual text.
    d. “Rolling Window” allows you to look at word trends across **ONE** document.
5. Analyze
    a. The “Analyze” tab provides a drop down menu of more analytical options for analyzing your texts.
    b. This includes comparing texts through functions like “Top Words” and “Similarity Query.”
6. To start over, select the Reset tab. It will clear the previously uploaded documents.
7. This is just a sample of the wide variety of tools Lexos offers for exploration.

Developed by: Colleen Nugent and Sayyara Poliferno
Questions? Contact us: nulab.info@gmail.com