



---

**HIST 2430 Digital Histories of Ethnic Boston**  
**Simon Rabinovitch**  
**Text Analysis using Lexos**

---

**Important Terminology**

- Corpus/corpora: a text or collection of texts that is used for analysis
- Stopwords: the most common words in a language
  - (English: the, a, she, her, it, they, etc.)

**Lexos website:** <http://lexos.wheatoncollege.edu/>

**What is Lexos?**

Lexos is a powerful web-based tool to help explore different corpora of digitized texts. Lexos provides a workflow of effective practices to help you be cognizant of choices you are making. For Lexos, separate .txt files are required.

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

**Step-by-Step Guide for using Lexos:**

1. Upload
  - a. Upload your corpus (your separate .txt. files)
  - 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 preparations, click “Apply” and wait a few minutes
  - c. You can also download your corpus after preparation.



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. This is just a sample of the wide variety of tools Lexos offers for exploration