Lexos

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Lexos: http://lexos.wheatoncollege.edu/upload

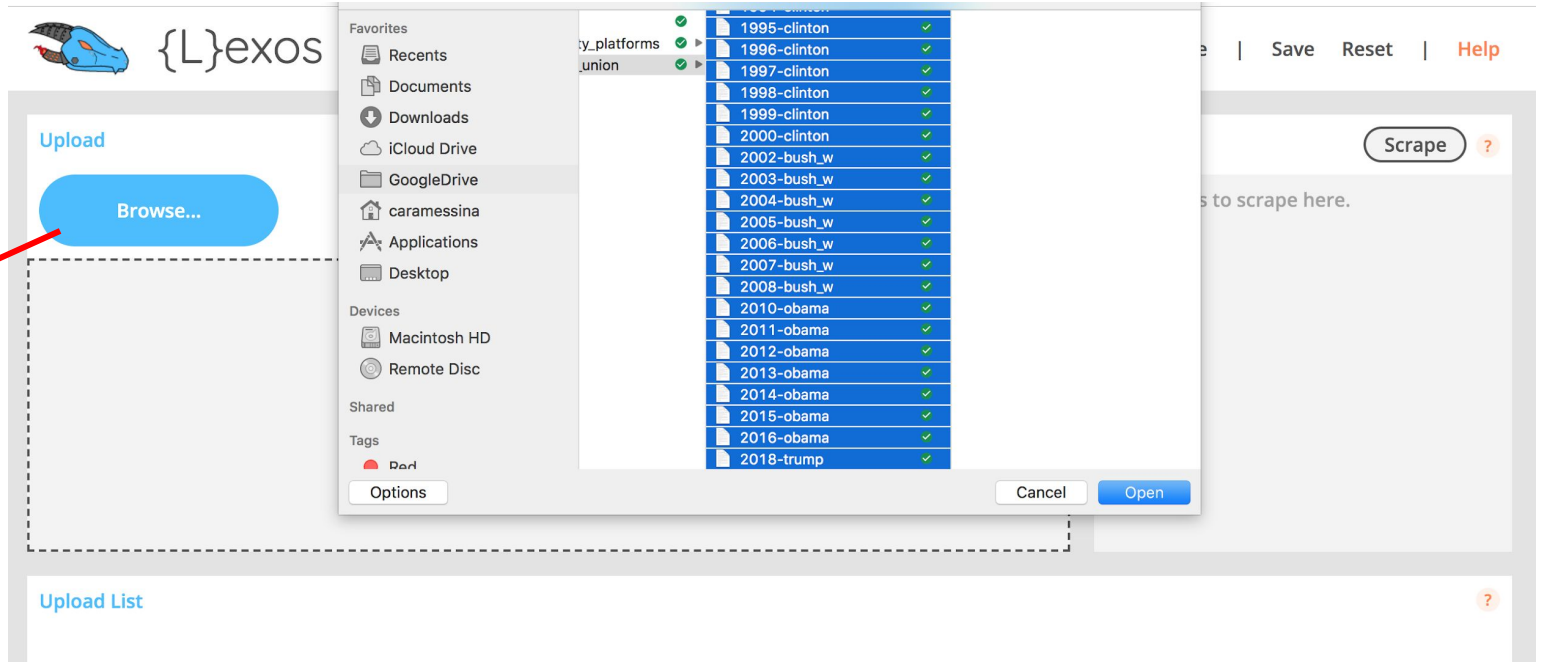
Lexos provides a step-by-step guide for corpus uploading, preparation, and analysis.

- **Upload**: upload your corpus (your separate .txt files)
- **Manage**: select the files you want to prepare and analyze
- **Prepare**: prepare your corpus for analysis
- **Visualize**: create visualizations of patterns across your corpus or in single texts
- **Analyze**: analyze your corpus, including comparing texts

Feel free to ask questions at any point during the presentation!

Lexos: Upload

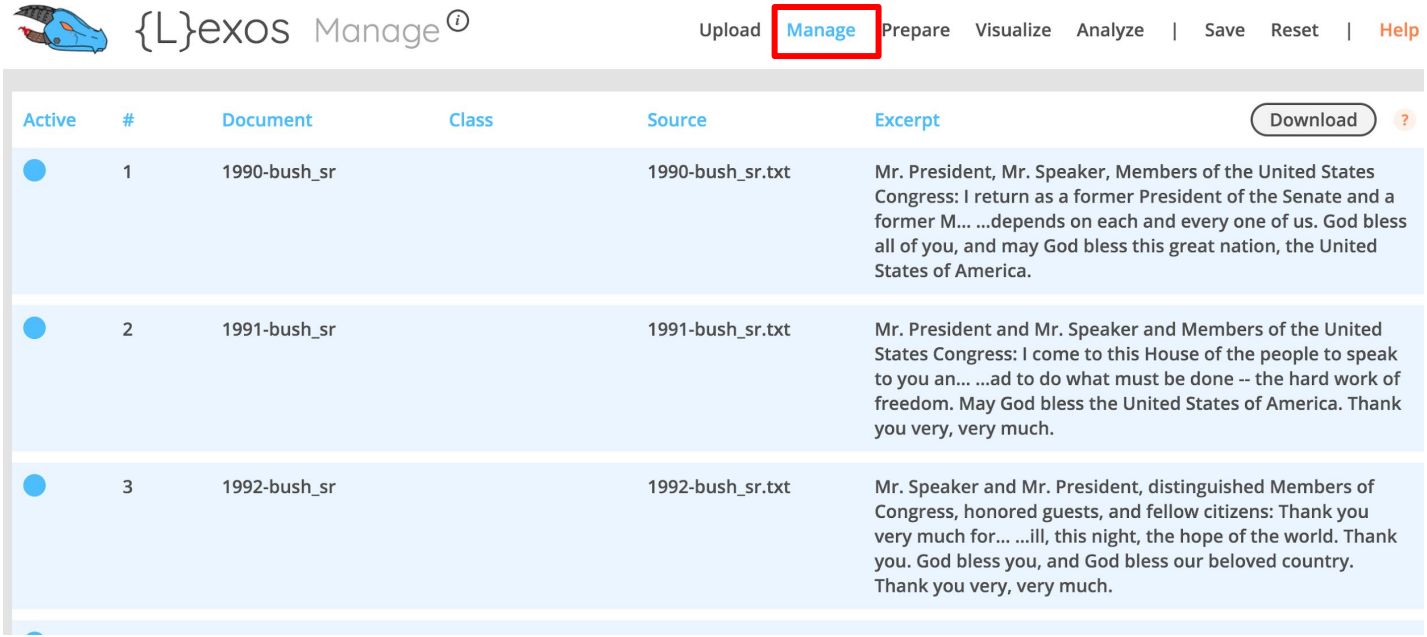
Click Browse and select your entire corpus (or drag and drop)



Feel free to ask questions at any point during the presentation!

Lexos: Manage

Make sure all the documents in the corpus you want to use are selected (blue = selected, gray = not selected)



Feel free to ask questions at any point during the presentation!

Lexos: Prepare (scrub)

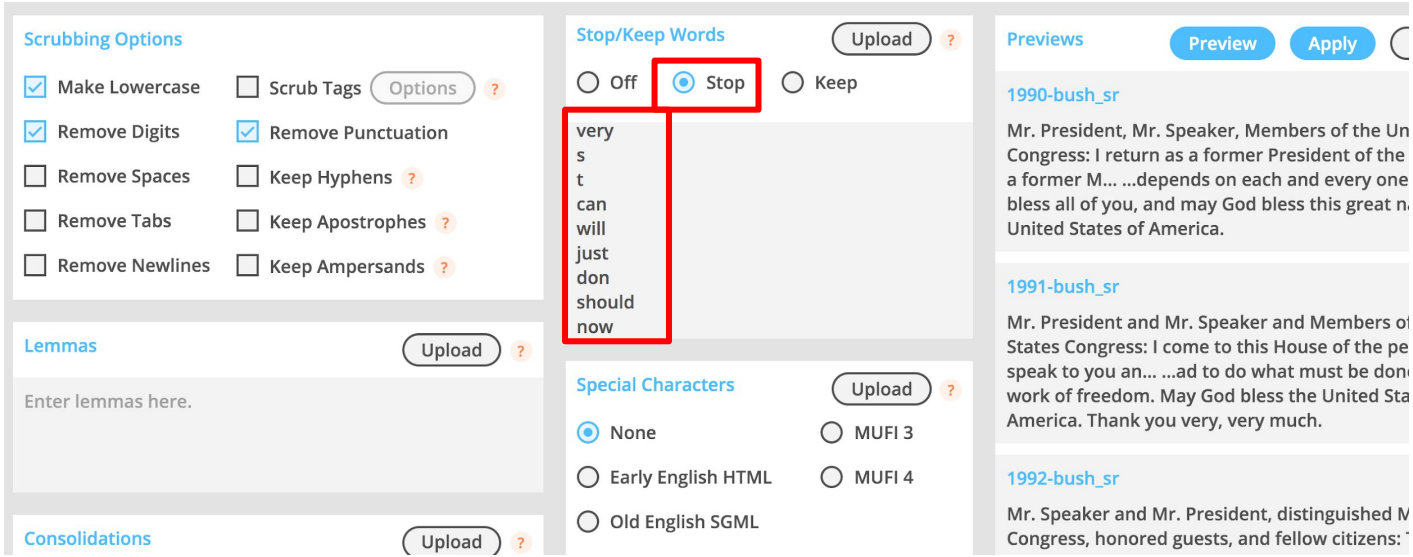
Lexos demonstrates the different options you have for preparing your corpus. By "scrubbing," you are transforming the texts in your corpus and making choices that will impact your results. Here are some possibilities:

- **Make Lowercase**: make all your letters lowercase. Even though you know "A" and "a" are the same letter, the computer treats these as two separate characters. Lowercasing removes this distinction.
- **Remove Punctuation**: remove punctuation, which may influence your results.
- **Stop/Keep Words**: remove a list of words. Usually these would be **stopwords**, or the most common words in a language (English: the, a she, her, it, him, they, etc).
- **Lemmas**: standardize to the *stem* of word. For example, you can stem all forms of talk: talking, talked, talks, etc. to "talk"

Feel free to ask questions at any point during the presentation!

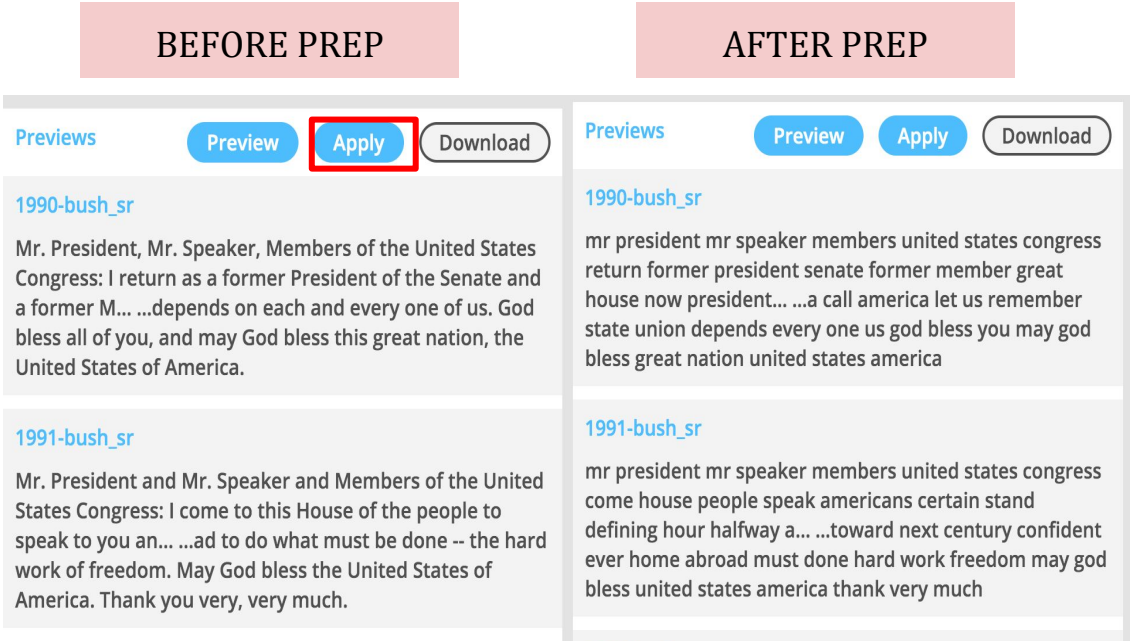
Lexos: Removing Stopwords

Get a list of English stopwords here: https://gist.github.com/sebleier/554280 (there is also a copy on the GitHub page). Copy and paste the stopwords (or upload the .txt file) into the "Stop/Keep Words" box then select "Stop"



Lexos: Applying your Preparations

Once you have made decisions about your preparations, click "**Apply**" and wait a few minutes. Because the program is going through each document and completing all the processes you selected, it needs some time. Then, you will see the final results of your preparation! You can also **download** your new corpus.



Feel free to ask questions at any point during the presentation!

Lexos: Visualize

Word Cloud: visualize a wordcloud across the entire corpus.

Multi Cloud: visualize wordclouds for each individual document/text

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Lexos: Rolling Window

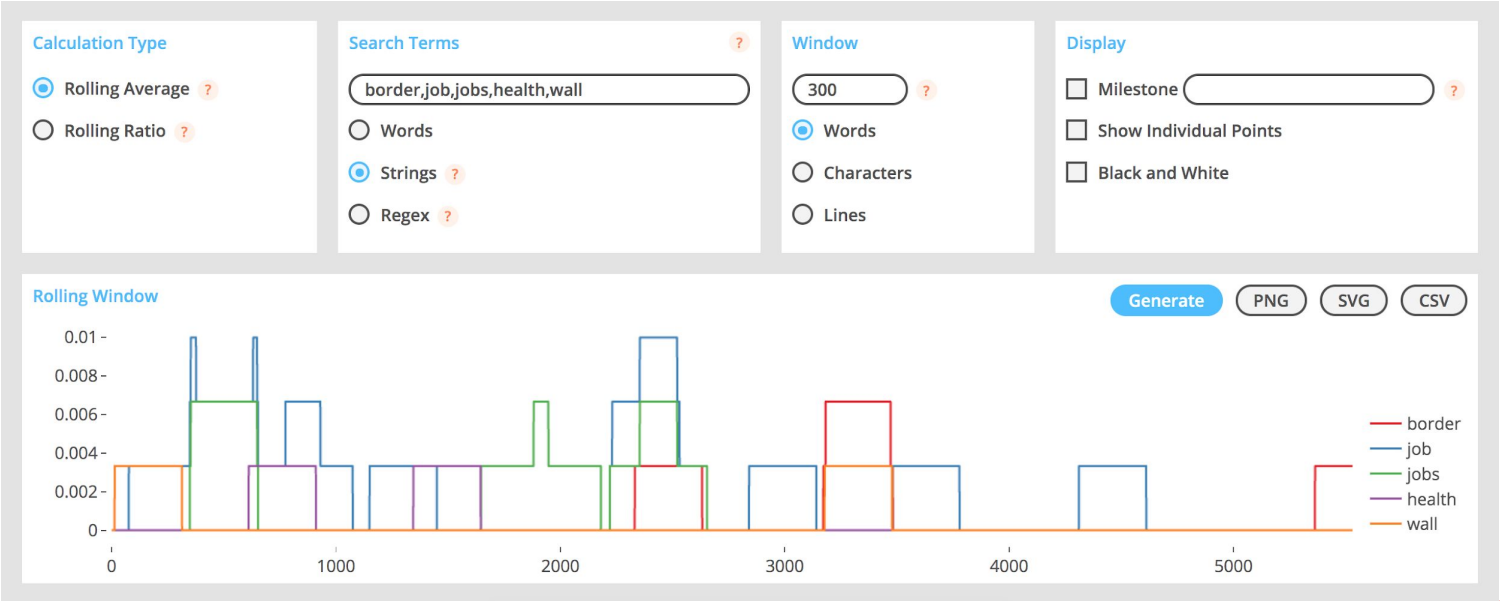
Rolling windows allow you to look at word trends across **one** document. To use a rolling window:

1. Go to “Manage” and right click one blue dot. Click “Deactivate all”
2. Choose the one **document** you want to analyze with the rolling window
3. Go back to “Visualize-> Rolling Window” and type in a search term you want to visualize. You can also search multiple terms by clicking “String” and separating words with a comma (president,health,republican)
4. Choose a Window size (the number of words each “window” contains). For shorter documents, it’s good to have a number like 300/500. For larger documents, you may want to make your window larger. Play around with the window size until you get a visualization that makes sense.
5. Click “Generate”

Feel free to ask questions at any point during the presentation!

Lexos: Rolling Window Results

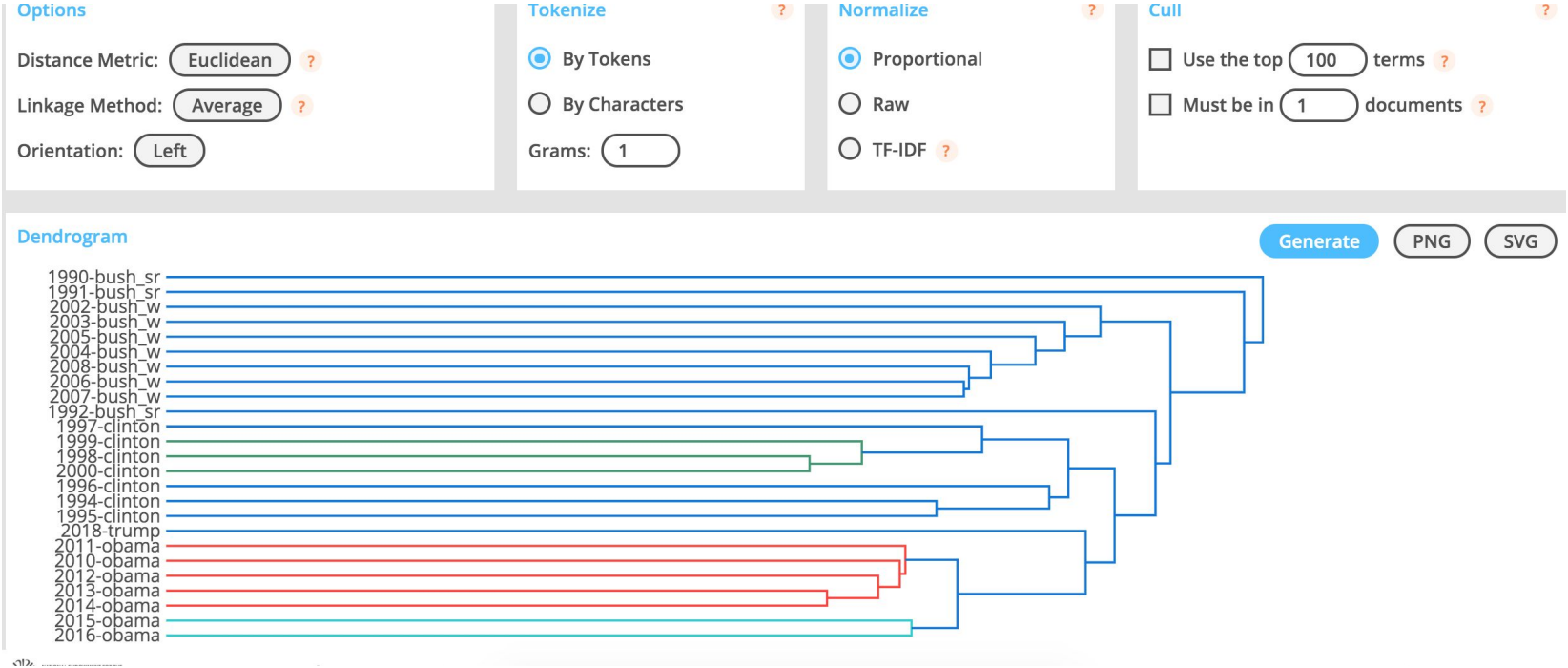
Using Trump's State of the Union from 2017, and searching for the strings "border,job,jobs" with a window of 300, we can get an idea of how different terms work together in Trump's speech. You may also be interested in **contrasting** terms to see how they're used across a text.



Feel free to ask questions at any point during the presentation!

Lexos: Analyze, Dendrogram

The dendrogram demonstrates similarity between the different documents.



Lexos: Save or Reset Your Results

Lexos allows you to **save** your results as a Lexos file. If you do this, you can re-upload the Lexos file any time to access your cleaned-up corpus as well as the different analyses you've done.

You can also save individual visualizations as images (PNGs).

Finally, if you want to start over, you can "Reset" your Lexos dashboard.

Feel free to ask questions at any point during the presentation!