

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>

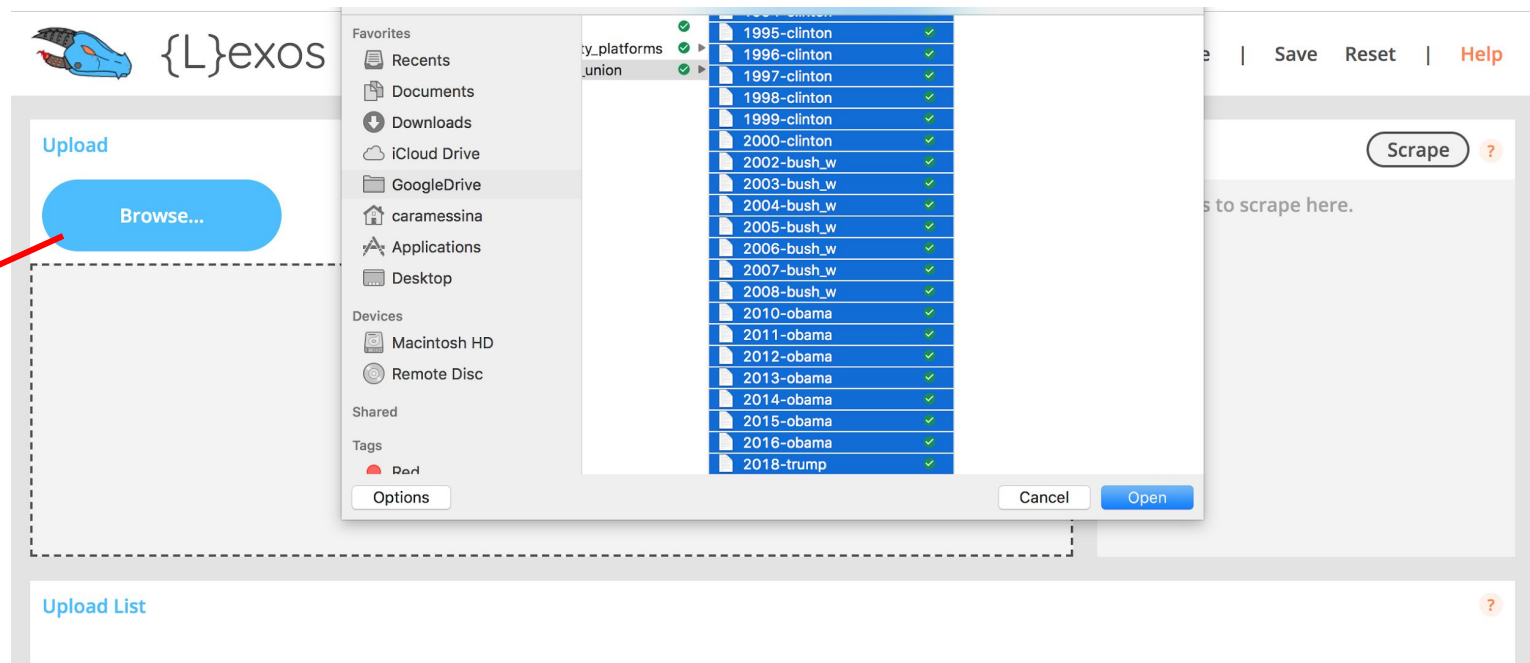
Lexos provides a step-by-step guide for corpus uploading, preparation, and analysis.

- **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



Lexos: Upload

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



Lexos: Manage

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



The image shows the Lexos Manage web interface. At the top, there is a navigation bar with the Lexos logo (a blue dragon head) and the text "{L}exos Manage". To the right of the logo are several buttons: "Upload", "Manage" (highlighted with a red box), "Prepare", "Visualize", "Analyze", "Save", "Reset", and "Help". Below the navigation bar is a table with the following columns: "Active", "#", "Document", "Class", "Source", "Excerpt", and a "Download" button with a help icon. The table contains three rows of data, all of which are selected (indicated by blue circles in the "Active" column).

Active	#	Document	Class	Source	Excerpt	Download
<input checked="" type="radio"/>	1	1990-bush_sr		1990-bush_sr.txt	Mr. President, Mr. Speaker, Members of the United States Congress: I return as a former President of the Senate and a former M... ...depends on each and every one of us. God bless all of you, and may God bless this great nation, the United States of America.	
<input checked="" type="radio"/>	2	1991-bush_sr		1991-bush_sr.txt	Mr. President and Mr. Speaker and Members of the United States Congress: I come to this House of the people to speak to you an... ...ad to do what must be done -- the hard work of freedom. May God bless the United States of America. Thank you very, very much.	
<input checked="" type="radio"/>	3	1992-bush_sr		1992-bush_sr.txt	Mr. Speaker and Mr. President, distinguished Members of Congress, honored guests, and fellow citizens: Thank you very much for... ...ill, this night, the hope of the world. Thank you. God bless you, and God bless our beloved country. Thank you very, very much.	



Lexos: Prepare (scrub)

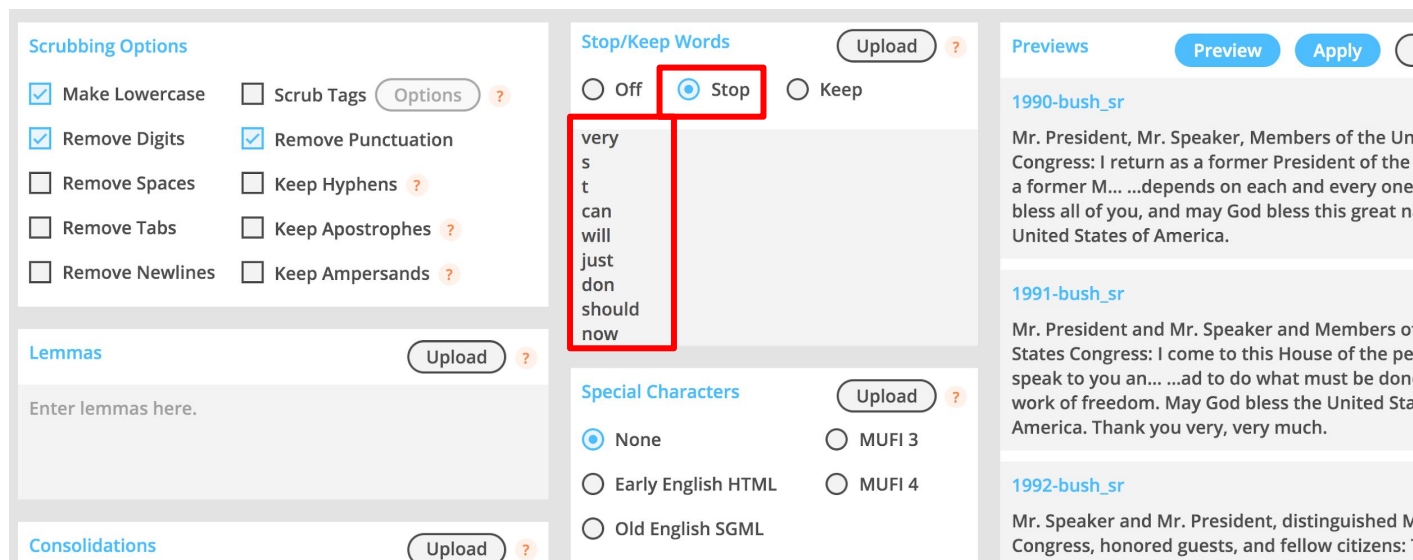
Lexos demonstrates the different options you have for preparing your corpus. By “scrubbing,” you are transforming the texts in your corpus and making choices that will impact your results. Here are some possibilities:

- **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”



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”



The screenshot displays the Lexos web interface with several sections:

- Scrubbing Options:** Includes checkboxes for 'Make Lowercase', 'Remove Digits', 'Remove Spaces', 'Remove Tabs', 'Remove Newlines', 'Scrub Tags', 'Remove Punctuation', 'Keep Hyphens', 'Keep Apostrophes', and 'Keep Ampersands'. Each option has an 'Options' button with a question mark.
- Lemmas:** A text input field labeled 'Enter lemmas here.' with an 'Upload' button and a question mark.
- Consolidations:** A section at the bottom with an 'Upload' button and a question mark.
- Stop/Keep Words:** This section is highlighted with a red box. It contains three radio buttons: 'Off', 'Stop' (which is selected), and 'Keep'. Below the radio buttons is a list of stopwords: 'very', 's', 't', 'can', 'will', 'just', 'don', 'should', and 'now'. The entire list is enclosed in a red box.
- Special Characters:** Includes radio buttons for 'None' (selected), 'Early English HTML', and 'Old English SGML', along with 'MUFI 3' and 'MUFI 4' options. Each option has an 'Upload' button and a question mark.
- Previews:** A section on the right showing three preview cards for '1990-bush_sr', '1991-bush_sr', and '1992-bush_sr'. Each card contains a snippet of text from a speech.



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.

BEFORE PREP	AFTER PREP
<p>Previews Preview Apply Download</p> <p>1990-bush_sr</p> <p>Mr. President, Mr. Speaker, Members of the United States Congress: I return as a former President of the Senate and a former M... ..depends on each and every one of us. God bless all of you, and may God bless this great nation, the United States of America.</p> <p>1991-bush_sr</p> <p>Mr. President and Mr. Speaker and Members of the United States Congress: I come to this House of the people to speak to you an... ..ad to do what must be done -- the hard work of freedom. May God bless the United States of America. Thank you very, very much.</p>	<p>Previews Preview Apply Download</p> <p>1990-bush_sr</p> <p>mr president mr speaker members united states congress return former president senate former member great house now president... ..a call america let us remember state union depends every one us god bless you may god bless great nation united states america</p> <p>1991-bush_sr</p> <p>mr president mr speaker members united states congress come house people speak americans certain stand defining hour halfway a... ..toward next century confident ever home abroad must done hard work freedom may god bless united states america thank very much</p>



Word Cloud:
visualize a
wordcloud across
the entire corpus.

A word cloud visualization of the text '1992-bush_sr'. The words are arranged in a circular pattern, with 'must' and 'know' being the largest and most prominent. Other visible words include 'use', 'america', 'billio', 'put', 'in', 'big', 'get', 'go', 'you', 'let', 'it', 'one', and 'ushelp'. The words are in various shades of blue and green.

A word cloud visualization of the text. The words are arranged in a circular pattern, with 'health' and 'care' being the largest and most central. Other prominent words include 'need', 'many', 'tell', 'get', 'every', 'year', 'know', 'us', 'time', 'welfare', 'job', 'continue', 'one', 'give', 'bill', 'cut', 'way', 'law', 'war', 'ISO', and 'PNG'. The words are in various shades of blue and green, and some are enclosed in rounded rectangular boxes.

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”



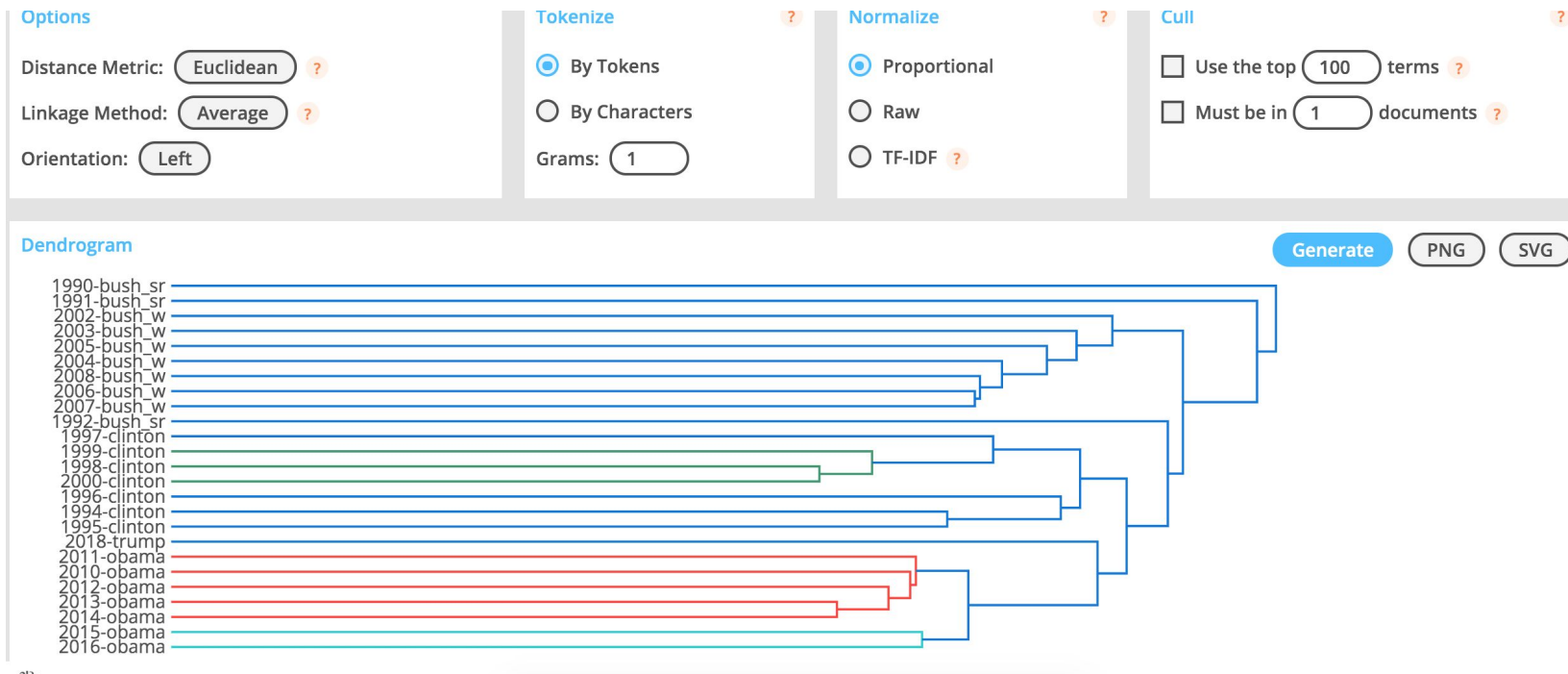
Lexos: Rolling Window Results

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.



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.

