



ENGL 4710 Capstone Seminar
Prof. Mya Poe
Text Analysis using Voyant Tools

Important Terminology

- Corpus/corpora: a text or collection of texts that is used for analysis.
 - Example: One could create corpus of all of Barack Obama's presidential speeches to trace his language over the course of his presidency
- nGram: A continuous sequence of n items in a text. For example, in Barack Obama's speeches, a bigram (or 2 continuous words) could be 'United States,' while a tripgram (3 words) could be 'yes we can.'

Voyant Tools website: <https://voyant-tools.org/>


What is Voyant?

Voyant is a powerful web-based text analysis platform "designed to facilitate reading and interpretive practices for digital humanities students and scholars as well as for the general public.". With Voyant, you can analyze one or more text files and use multiple visualization tools

Voyant can read .pdf and .doc, but .txt is recommended.

For Voyant documentation see: <https://voyant-tools.org/docs/#!/guide/about>

Step-by-Step Guide for using Voyant:

1. On the Voyant front page, click on 'upload' and navigate to the chosen corpus to upload text files - then click on 'reveal'.
 - a. Alternatively paste URLs or full text into textbox
 - b. Advanced options and help can be found in the top right of the text box.
2. The results are given in five default panes that can be changed.
3. To change a pane:
 - a. Hover mouse over one of the panes.
 - b. Click the panes button: 
 - c. Use the dropdown menu to navigate to different tools.

Slides, handouts, and data available at <https://bit.ly/fa22-poe-textanalysis>

Developed by: Jonathan Sullivan & Colleen Nugent, DITI Research Fellows

Questions? Contact us: nulab@northeastern.edu



A few notes:

- Your results will show a fairly busy space with a number of different tools for examining your input document(s). In our own analyses, we'll focus on just a couple of these, but you should feel free to experiment with any that seem potentially useful to you.
- Voyant provides a range of tools extending well beyond what is in the default; you can see a full list of the options if you click on the icon that looks like a window in the top-right of the screen.
- Voyant provides some basic help and documentation if you click on the question marks for each tool.

Results:

- The “Trends” tool (which appears in the top-right corner in the default Voyant results display) allows you to examine the usage for a word or set of words across a text or corpus. It will default to showing the most frequent (non-stopwords) terms in your input document(s), but you can fill in any terms you would like to query in the search box. There are fairly sophisticated searching options, which you can see by mousing over the question mark next to the search box. You can experiment with the settings using the gear icon.
- The “Contexts” tool (which appears in the bottom-right corner in the default results) allows you to examine keywords in context. You can input query terms in the search box (again, with many different search options) and use the arrows to sort by the contexts to the left or right of your input term. This is far from the most sophisticated or useful way to examine keywords in context—if you'd like to go further with this, check out [AntConc](#)—but it at least gives you a sense of how this approach works and why it might be useful.