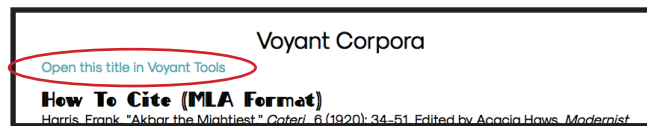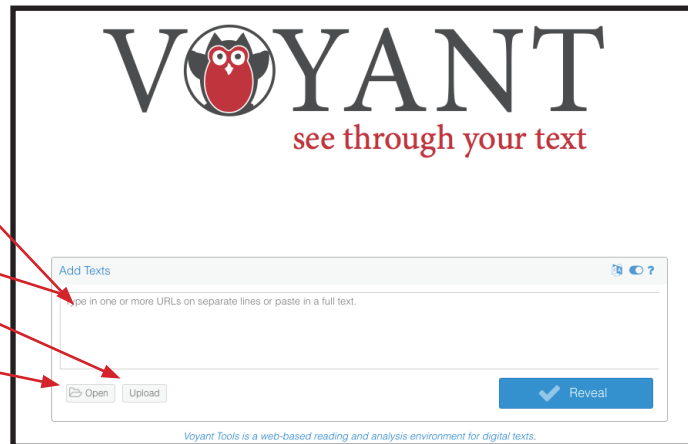# VOYANT TOOLS

Getting Started

Option 1: Analyze One Text from MSSP

Find the short story you want to analyze on the MSSP site, scroll down to the bottom of the page, and click the "Open this title in Voyant Tools" button

Voyant Corpora

Open this title in Voyant Tools

How To Cite (MLA Format)

Harris, Frank. "Akbar the Mightiest." *Cateri*, 6 (1920): 34–51. Edited by Acacia Haws. *Modernist*

Option 2: Make Your Own Corpus (for analyzing 2+ texts or texts outside of MSSP)

1. Go to www.voyant-tools.org

2. Submit plain text version of your file. There are a couple ways to do this:

    a) Upload a link to an online text file - .txt (best on MSSP, Project Gutenberg doesn't do this well)

    b) Copy and paste text directly into the box

    c) Upload a text file from your computer (.txt file comes from a text editing app like TextWrangler)

    d) Open an existing corpus from Voyant (Austen or Shakespeare)

4. Click Reveal.

VOYANT *see through your text*

Add Texts

Type in one or more URLs on separate lines or paste in a full text.

Open Upload   Reveal

*Voyant Tools is a web-based reading and analysis environment for digital texts.*

Default Dashboard

Voyant has many tools, but automatically populates with five tools: Cirrus, Reader, Trends, Summary, and Contexts. All of the sections are automatically populated with the most frequent word(s). You can analyze any words and narrow specific documents with the settings at the bottom. The sections interact with each other. (e.x., if you click on a word in the Cirrus cloud, that word will be highlighted in the Reader and graphed in the Trends.)

The "Cirrus" section displays word frequency. The larger the word, the more frequently it occurs.

The "Reader" section allows you to scroll through the entire uploaded text.

The "Summary" section shows details about the documents included in the corpus.
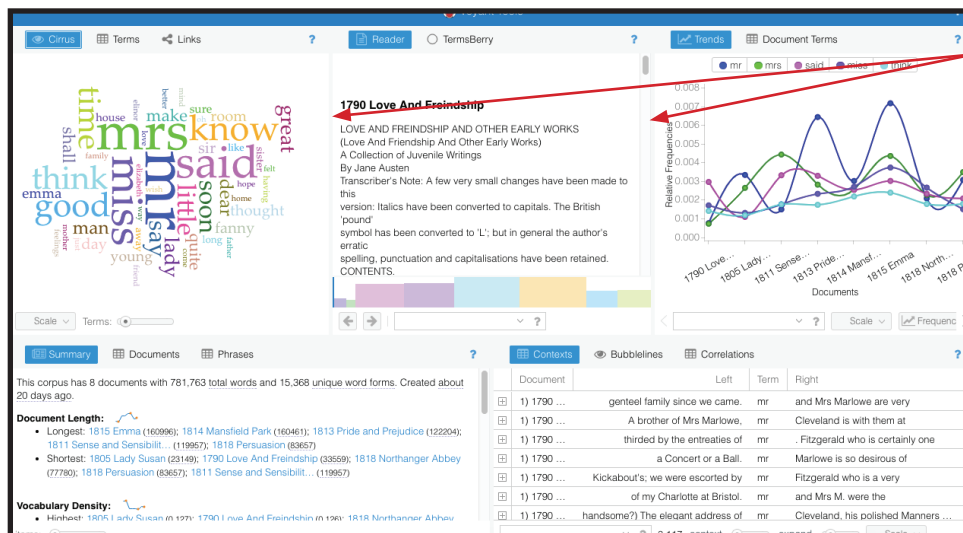
All of the section window sizes are adjustable for your needs.

"Trends" shows word frequency in each text or each section of a single text.

The "Context" section shows the selected word w(or most popular word) with its surrounding words.
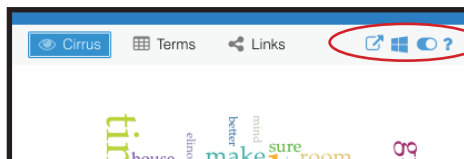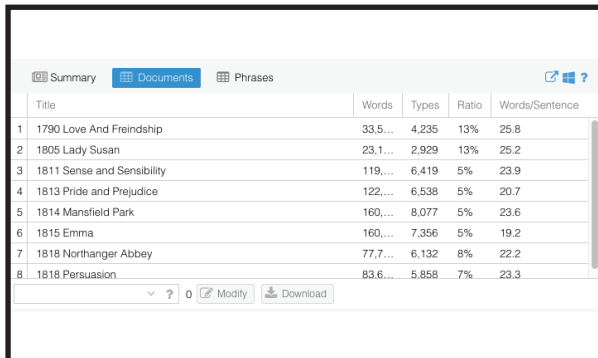
Hover on the icons in the top right corner of each section to view more options. The **box with an arrow** exports a link to the analysis tool (or the entire corpus if you use the icon in the top right of the whole page). The **cluster of boxes** shows a dropdown menu of the many analysis tools available in Voyant. Any tool can be shown in any section, so Voyant is entirely customizable for your needs. The **switch icon** opens a window for more options to customize the tool in use. The **question mark** opens a separate Help page for Voyant.

Tips

Search Terms

You aren't limited to the words that the default system runs. You can actually search the frequency of any word in the text you are analyzing with most of the tools on Voyant using the dropdown/search bar at the bottom. If you click the down arrow on the search bar, the ten most frequent words will appear. Or, if you click inside the search bar, you can type any word. There are also some search tricks you can use to make sure you're catching all the terms you're looking for. If you hover your cursor over the question mark, these tips will appear. They are also listed below.

- `^coat*`: match terms that start with coat as **separate terms** (coat, coats, etc.)
- `*oat`: match terms that end with the **suffix** oat as **one term**
- `^*oat`: match terms that end with **suffix** oat as **separate terms** (coat, moat, etc.)
- `coat,jacket`: match each term **separated by commas** as **separate terms**
- `coat|jacket`: match terms **separated by pipes** as a **single term**
- `"winter coat"`: winter coat as an exact **phrase** (word order matters)
- `"winter coat"~0`: winter coat or coat winter **phrase** (word order doesn't matter but
- `"coat mittens"~5`: match coat **near** mittens (within 5 words)
- `^coat*,jacket|parka,"coat mittens"~5`: **combine** syntaxes

Palette

For visual analysis tools like Cirrus and Knots, you can choose your color palette. Click on the settings button (the switch icon in the top left corner of the tool box) and in the pop-up box, click the "Edit List" button. From here, you can add/remove colors, and edit the existing colors.

Stop Words

"Stop Words" are the words that are excluded in an analysis. The default setting on Voyant removes stop words (like pronouns, prepositions, article adjectives) so you have a more accurate view of what meaningful words are used.

    A) no stop words are excluded from this analysis
    B) the default analysis, removes auto-detected English stop words
    C) the auto-detected English stop words are removed, along with "mr," "mrs," and "miss"

You can customize the analysis in the settings (the switch icon in the top right corner) and either add or remove stop words with the "Edit List" button (D & E)

VOYANT TOOLS

Tools - Basics

Documents

The Documents tool shows a table of the documents in the corpus and allows you to add/remove documents from the corpus with the "Modify" button at the bottom.

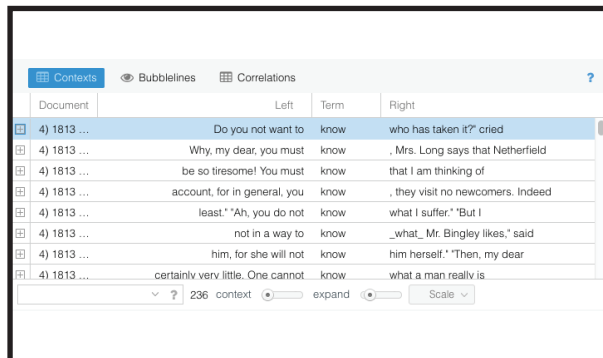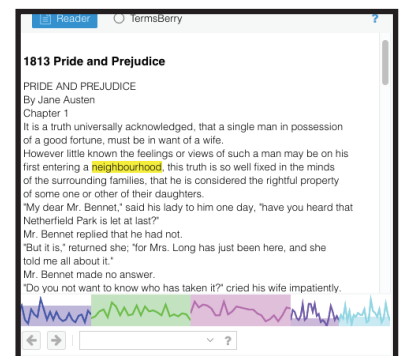*Words*: # of total words in the doc (e.g. each occurrence of "the" is counted individually)
Types: # of unique words in the doc (e.g. all occurrences of "the" count as one word form)
Ratio: the ratio of unique words to total words, expressed as a percentage. (higher numbers generally mean greater vocabulary diversity)
Words/Sentence: the number of words per sentence (approximated)

Reader

The Reader tool provides a way of reading documents in the corpus. Hovering over a word will reveal its frequency in the document. If you're analyzing several texts, they will be divided into separate colors at the bottom. Clicking on a word will highlight all of its appearances in the text, and will graph the frequency of the word in the bottom bar.

Contexts

The Contexts tool shows each occurrence of a keyword with a bit of surrounding text. It can be useful for studying more closely how terms are used within the context of the text or texts analyzed. You can search for specific terms in the bottom search bar. If you want to see the larger context, click the "+" icon on the left. If analyzing multiple texts, you can customize which texts are included with the "scale" option at the bottom.

Cirrus

Cirrus is a word cloud, displaying the words that appear most frequently in the text you're analyzing. The more frequent the word, the larger it appears. If you hover your mouse over the word, you can see the number of times it appears in the text or corpus. You can use the "Terms" slider at the bottom to include more or fewer words in the cloud. The Cirrus tool is a good starting point for generating questions about the text you want to analyze.
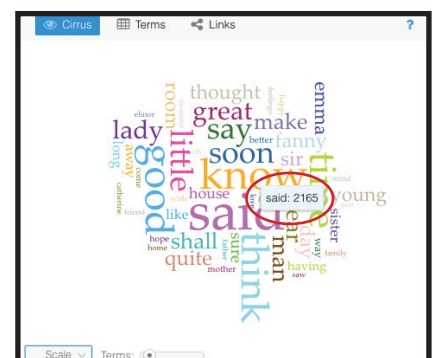
VOYANT TOOLS

Tools - Basics
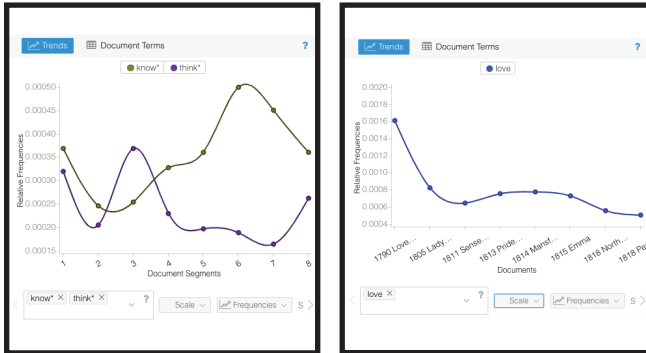
Trends

Trends shows a line graph depicting the occurences of one or more words across a corpus or document. Use the search box at the bottom to view specific words (the default is the most frequent words in the text). If analyzing only one text, you can customize the number of segments into which the text is divided. Raw frequencies show the total number of occurences, while relative frequencies take into account the length of the text (generally more accurate).

Collocates

Collocates is a table view of the keywords and collocates (terms that occur in close proximity) in the analyzed text(s). You can search your own keywords in the bottom search bar and use the context slider to adjust the number of collocates included on each side of the keyword.
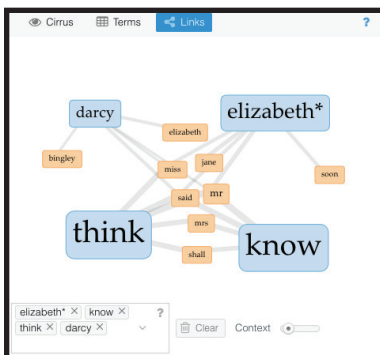
Term: this is the keyword (or keywords) being searched
Collocate: these are the words found in proximity of each keyword
Count (context): this is the frequency of the collocate occurring in proximity to the keyword

Links

Links shows keywords and collocates (terms that occur in close proximity) as a force directed network graph with keywords in blue and collocates in orange. Hovering over a word shows its frequency. You can drag words to move them. You can drag words out of the box to remove them. Search for your own keywords in the bottom search bar. The context slider adjusts the number of collocates included on each side of the keyword.

Mandala

Mandala shows the relationships between keywords and documents in a corpus. Each keyword pulls documents toward it based on the term's relative frequency in the corpus. You can remove default keywords by clicking on a word and clicking "remove" in the pop up box. You can add keywords by clicking "add" and typing in a keyword—but you can only add one at a time.
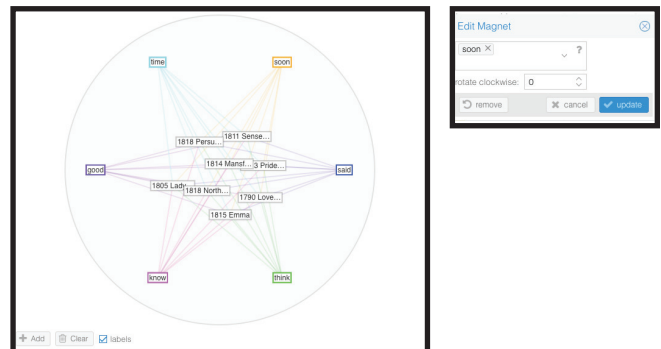
Other Tools

These are a few of the 27 analysis tools available on Voyant. For further explanations on those tools and the tools listed above, click on the "Help" icon within Voyant (the question mark icon). This section of the Voyant program also contains a "Getting Started" guide and a "Gallery" of example projects.

Feel free to play around with the different tools available to find the best one to fit your analysis project.

Voyant Help: http://voyant-tools.org/docs/#!/guide

List of Voyant Tools:
- Bubblelines: visualization of the frequency and distribution of terms in a corpus
- Bubbles: visualization of term frequencies by document in different-sized bubbles.
- Catalogue: interface for exploring a larger, multi-document corpus based on search criteria and filters
- Cirrus: a word cloud that shows top frequency words of a corpus or document
- Collocates Graph (Links): representation of keywords and and collocates (words in close proximity) in a network graph
- Corpus Collocates (Collocates): a table view of which words appear most frequently in close proximity to keywords
- Contexts: table view of each occurence of a keyword with the surrounding context
- Corpus Terms (Terms): table view & line graph of keyword frequencies in the full corpus
- Correlations: line graph visualization showing whether the frequency of two keywords occur in sync or inversely
- Documents: table view of detailed info about documents in the corpus. Use this tool to modify your corpus
- Document Terms: table view & line graph of keyword frequencies in a document
- Knots: visualization that represents keywords in a document as a series of twisted lines
- Mandala: visualization showing the relationship between keywords and documents
- MicroSearch: visualization of frequency & distribution of keywords in a corpus in extra small text
- Phrases: table view of frequent sequences of words (by phrase frequency, repetition numbers, or number of words in phrase)
- Reader: interface for reading documents in the corpus and tracking keyword frequency
- Rezoviz: representation of connections between people, places, and organizations that co-occur in multiple documents
- Scatterplot: a more complex graph visualization, optimizing several data analyses in a multidimensional space
- StreamGraph: graph visualization of the change of frequency of keywords in a document or corpus
- Summary: detailed overview of iinformation about the documents in the corpus
- TermsBerry: visualization of high-frequency keywords and their collocates
- TermsRadio: visualization depicting the change of keyword frequecies in a corpus or document
- Textual Arc: weighted visualization of keywords in a document on an arc that follows the terms in document order
- Topics: rudimentary formation of word clusters (topics) from a document or corpus and distribution of those topics across document or corpus
- Trends: line graph depicting the distribution of a word's occurrence across a corpus or document
- Veliza: a (very) experimental tool for having a (limited) natural language exchange based on your corpus
- WordTree: visualization of how keywords are used in different phrases in the corpus