Narratives in News Media
Text Analysis with Voyant

Ryan Clement
March 18, 2019

Introduction

In working with text documents like the ones you have for your project, you may be familiar with the process of close reading – that is, detailed critical analysis of the language, structure, or patterns of a text. However, as you begin to build up your collection of sources, it will become more and more difficult to perform a close reading analysis on them, particularly as a group. Computers give us another option – we can engage in distant reading of these documents – not studying individual texts, but aggregating and analyzing collections of texts.

Many times, distant reading is done using programming languages, such as Python or R. Coding your own analyses offers an almost limitless degree of customization and granularity, but they have a bit of a learning curve. However, there are tools that have been built to aid distant reading, trading a bit of customization for usability. We’ll look at one of these tools today, Voyant (Sinclair and Rockwell n.d.).

Voyant Tools

Voyant, located at https://voyant-tools.org/, is a tool suite built by Stéfan Sinclair and Geoffery Rockwell, as part of the development of their website Hermeneuti.ca and book Hermeneutica. While you sacrifice some of the control and customization of coding your own analyses, Voyant provides a great path for getting started and doing light, exploratory analyses.

Loading Your Data

1. Navigate your browser to https://voyant-tools.org/ and click “Upload”
2. Find your documents. You can upload individual documents (PDFs, TXTs, etc.) or you can upload a ZIP file of all of your documents. You can also upload documents from a URL.
3. Once you’ve added all the documents, click the “Reveal” button to open the analysis screen

![Figure 1: The Voyant interface with the Shakespeare corpus loaded](image)

**Working with your data**

When you first open your documents in Voyant, take a look around the different windows. By default, you’ll have 5 tools on the screen when you open a set of documents:

1. **Cirrus**: a word cloud viewer
2. **Reader**: a place to read text from documents
3. **Trends**: line graphs showing the distribution of a word’s occurrence across a corpus or document
4. **Summary**: a simple, textual overview of the current corpus
5. **Contexts**: shows each occurrence of a keyword with a bit of surrounding text; can be useful for studying more closely how terms are used in different contexts

Some of the other tools will be available in the ‘tabs’ on the default tools. For example, in the same block as the Cirrus tool, we can switch to ‘Terms’ and ‘Links.’ Other tools can be added by clicking the little four-panel icon in the upper-right of each block. You can find information about all the various tools available at [https://voyant-tools.org/docs/#/guide/toolslist](https://voyant-tools.org/docs/#/guide/toolslist).

Each of the tools has the same set of menus in the upper-right corner. You can see them here:
The Export menu will let you export, as applicable: * a URL (that will let you ‘save’ your work) * an HTML snippet so you can embed your visualization in a web page * a citation for the tool * a PNG or SVG of the visualization itself

The Choose a New Tool menu lets you, well, choose a new tool.

The Options menu lets you define options for the current tool. These include: * stopwords * color palettes (for visualizations)

The Help Text menu gives you a brief overview of what the tools does.

Additionally, some tools also have options along the bottom of their block. Some important ones include the Scale menu, that lets you change whether a tool’s output is based on the full corpus or just a particular document (or anything in between), and the Display menu, that will let you set options for certain visualizations.

Notes on the major tools

Cirrus

Word clouds have a controversial reputation. They can be useful for quickly drawing attention to very high frequency terms. They have also been called “the mullets of the Internet” and have been criticized for being highly reductive and even misleading (Harris 2011).

Reader

Text is fetched for the Reader as you make selections in other tools. For instance, click on a document in the Documents viewer, or a line the Contexts tool, and see how the text in the Reader changes. The multi-colored bar along the bottom of the Reader is known as the ‘Prospect Viewer,’ which is a simple visual overview of the entire corpus. The taller/wider a bar is, the longer the document is. The thin blue vertical line shows your approximate position in the corpus.
Trends

Trends is a powerful visualization tool that represents frequencies of a term (or multiple terms) in a document (or multiple documents). You can set Trends to display data in multiple formats using the Display menu at the bottom of the block. You can also switch between two modes in Trends by double-clicking on of the dots that represents a term:

- **Terms** show the currently selected term for all documents
- **Document** for the current document, show all available terms (from the legend)

Don’t worry, if you mess up your visualization, you can always ‘Reset’ it to start all over again!

Summary

The Summary displays 6 categories of information formatted in a bulleted list:

1. An overview of the corpus
2. The top longest documents and shortest documents – the sparkline shows the documents in the order that they were added
3. The top vocabulary densities – the sparkline shows the documents in the order that they were added
4. An approximation of the average number of words per sentence, both the highest and lowest values.
5. The five most frequent words in the corpus
6. The top five most distinctive words of each of the documents

You can change the “Top 5” to be 10, 15, or whatever number of items you want using the “items” slider at the bottom of the block.

Contexts

Contexts shows terms in their context, giving a bit of text that comes before and after each terms displayed, as well as noting which document the context comes from. You can change the “amount” of context given by using the slider, and can change between the full corpus, a single document, or a group of selected documents using the Scale menu. Use the search box to look for particular terms, or click on them in one of the other tools to see their context in the Contexts tool.
References
