**Introduction**

Computer-assisted text analysis, or “text analytics,” supports a different kind of reading than we usually do in a literature course. Whereas in the latter we usually engage in “close reading,” when we do computer-assisted text analysis we are doing what some practitioners have come to call “distant reading” or “macroanalysis.” It involves stepping back from a text and, with the help of software, inspecting the language of the text as if it were any other kind of data. This approach allows us to look for patterns in the language that we might have missed otherwise.

Why would we want to do this? Because it encourages us to…

* revisit questions we thought we knew the answers to, and
* ask new questions that would never have occurred to us previously.

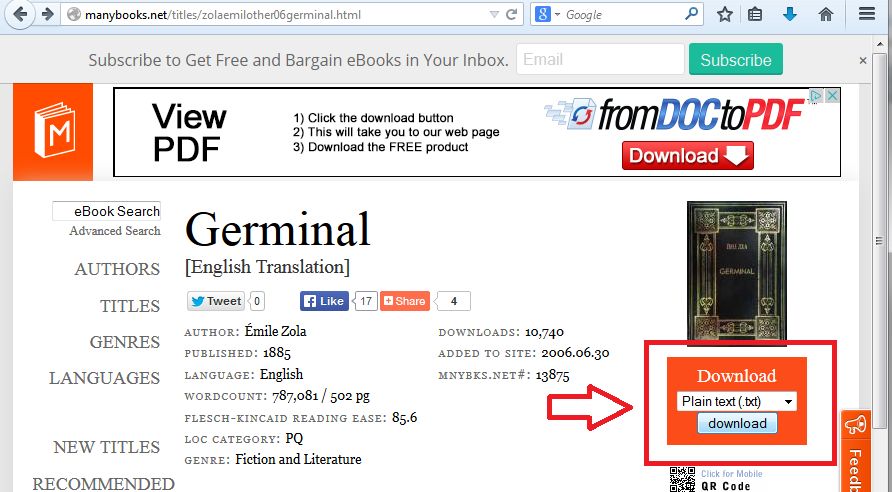
There are many web-based software tools that can help one perform text analytics. In this workshop, we’ll be using the Voyant Toolkit, a well-established suite of tools that offers a range of perspectives on the data of literary texts.

To complete this mini-workshop, you’ll need an internet connection and online copies of the texts we will be analyzing.

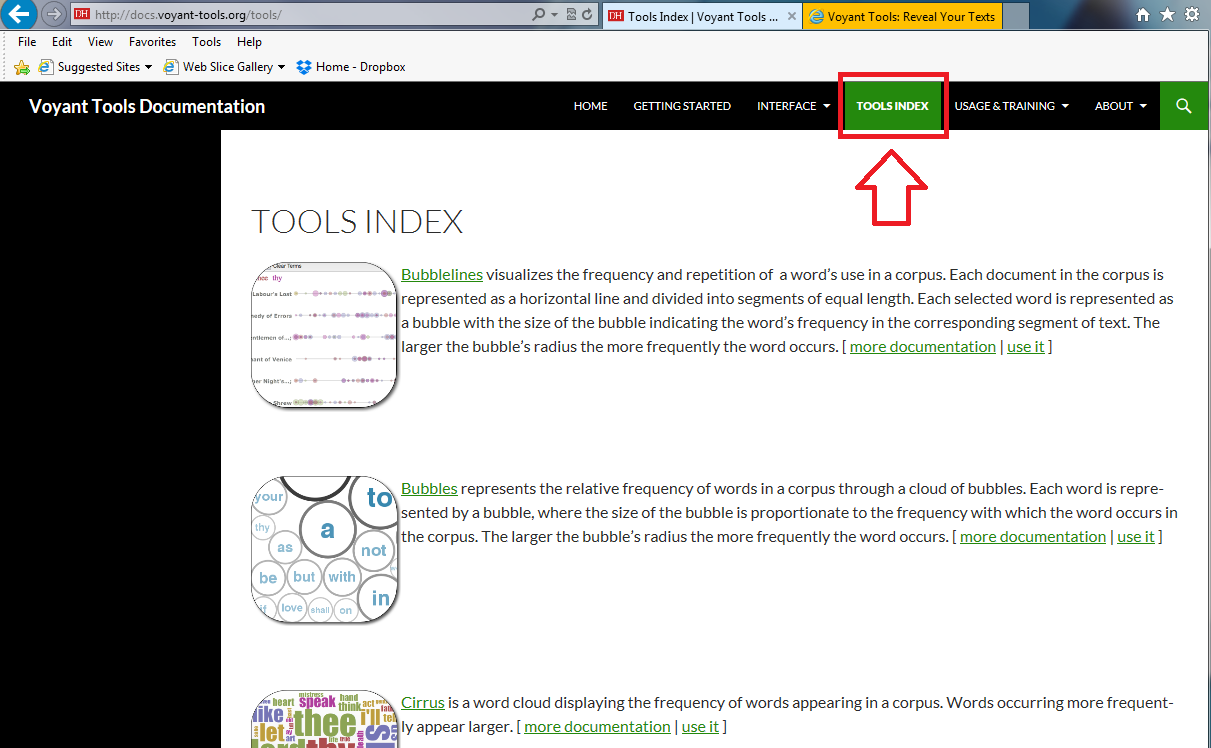
**How to do it**

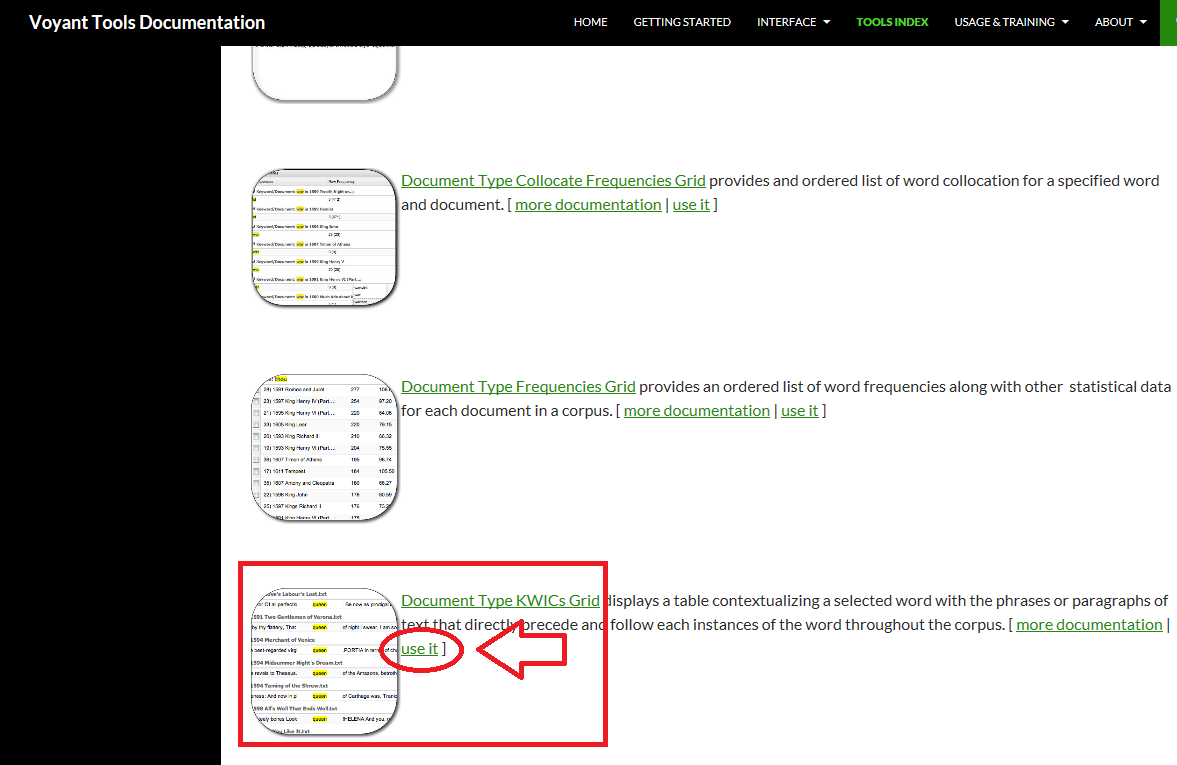
For this workshop you will use Voyant’s KWIC (Keyword in Context) tool to help you find and analyze similes in a literary text.

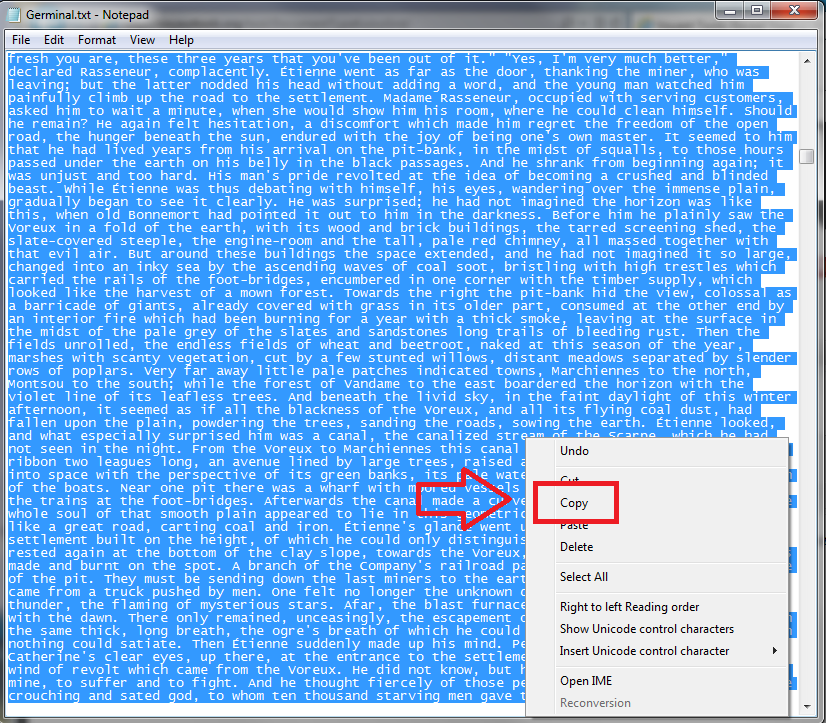
1. Download the target text (we are using a translation of Emile Zola’s *Germinal)* as a plain text file, located here: <http://www.gutenberg.org/ebooks/5711> . (Download the file labeled “Plain Text UTF-8”.)

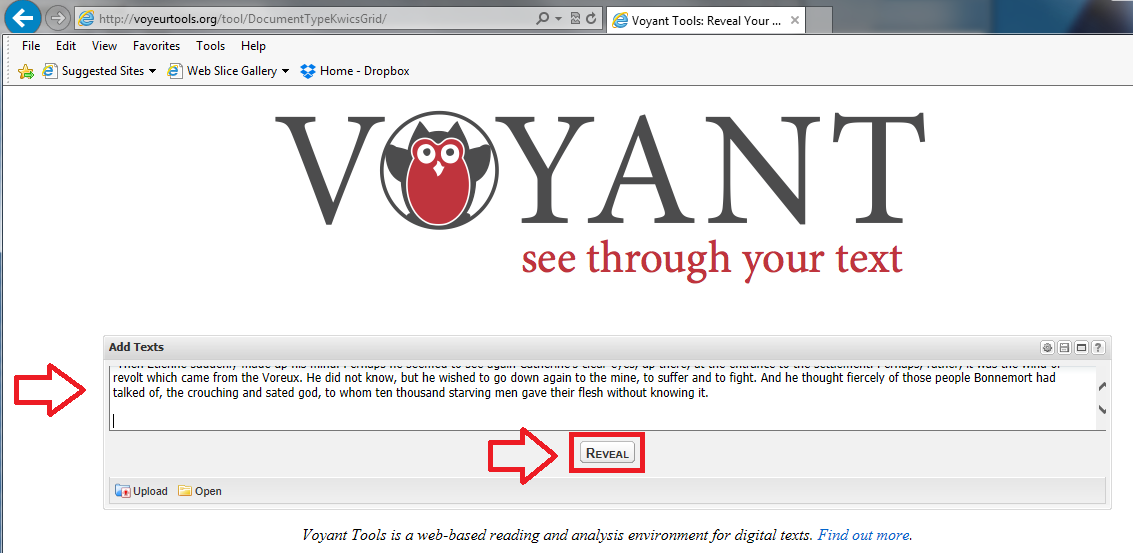


1. Go to <http://docs.voyant-tools.org/> and click **Tools Index**.

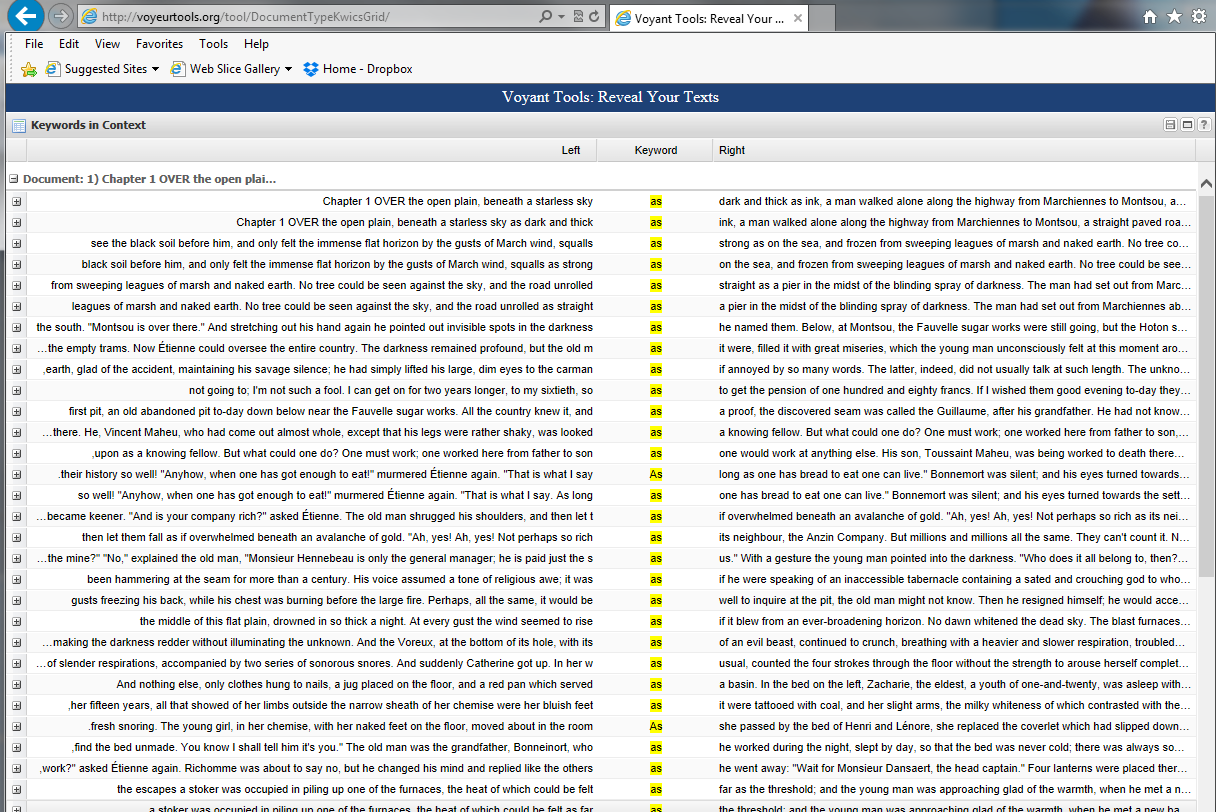
  
  
Page down until you see the tool called “Document Type KWICs Grid.” Click **use it**.

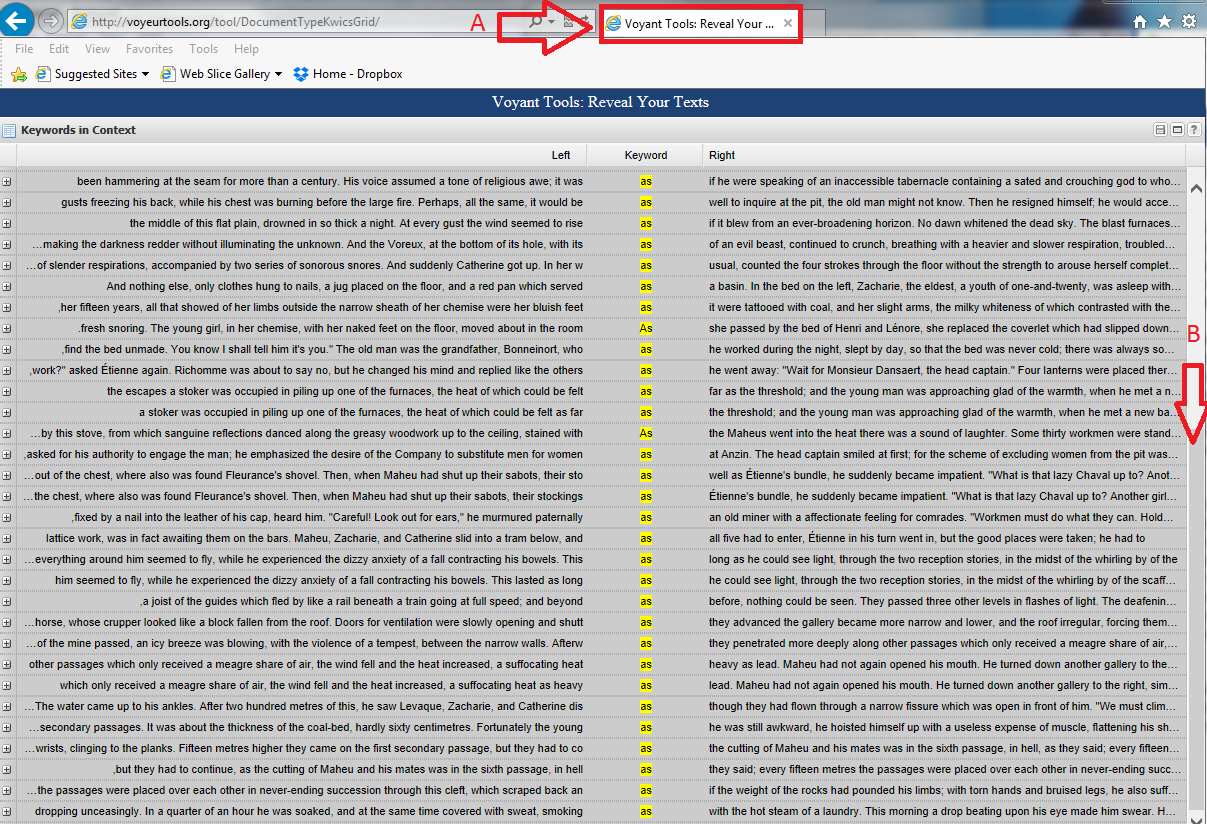


Copy and paste the target text into the Voyant **Add Texts** field, then click **Reveal**.  
   


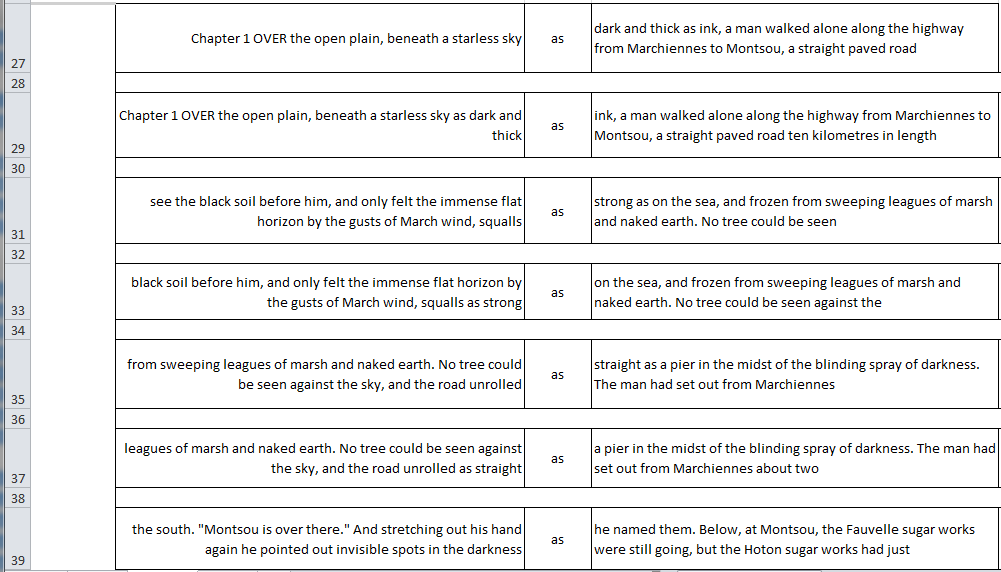


1. Now you’ll see a screen that says “No results.” That’s because this tool requires that you tell it specifically which “keywords in context” (“KWICs”) you want to see. Type the word “as” into the search field at bottom of the page. You should now see a three-column display that looks like this:   
     
   
2. Click the **Context** dropdown button at bottom of the page and select “20.”  
     
     
     
   This will expand to twenty words the context in which the term “as” appears in the display, which will allow you to tell at a glance if the word is or is not being used as part of a simile.



1. Now click on the top row of the KWIK listing. This will open a new window, which you will want to ignore. Instead, you will need to go back to the original KWIK listing by clicking the tab at top of the display (A). Then press and hold down the shift key and, using the slider on the far right, page down to the bottom row of the display (B).   
     
   
2. Once you’ve selected the entire page of KWIC data, click **Select all** and **Copy**.



1. Open an Excel or Google spreadsheet and paste the data into it, maintaining the formatting of the original. Every row from the first page of the KWIC tool display should appear in the spreadsheet. NOTE: You may have to tweak the row heights to make sure you can see all of the data that you want. If there are cells that got copied but that you don’t need, just delete (or ignore) them. You should see your keyword (“as”) and the ten or so words to the immediate left and right of the keyword, as shown here:   
     
   
2. Repeat the process for the rest of the pages that list the word “as” in context. Then do the same for “like” and other terms that may indicate a simile has been used. Once you’ve completed your inventory, remove any duplicates (there will be many in the “as” portion of the data, due to the common simile construction “as [ ] as a [ ]”). You’re now ready to look for patterns or themes in the similes.