

Distant Reading: *The Adventures of Sherlock Holmes*

Ann McClellan
Plymouth State University

Artifact type: Assignment

Developed for: Lower-level undergraduate seminar; also a general education “Quantitative Reasoning” course

Audience: English majors and minors

Time required: One 100-minute class session + outside work

Method and tool: Text analysis with Voyant

Description:

This assignment is designed for beginning students to build upon traditional textual analysis skills learned in literature and some general education courses. Students should already be familiar with the basic premises of close reading prior to the assignment. The motivation behind this specific project came from my university’s requirement for a Quantitative Reasoning connection in every major: “Mathematics finds application in all fields of scholarship. All disciplines make use of quantitative reasoning in some way and to some extent. Students take a three-/four-credit Quantitative Reasoning (Q) course specified as required for their major. This course may be taught within the major discipline or not. It might teach quantitative techniques used as primary or secondary tools within the discipline or might be a course in which students of less quantitative disciplines come to deepen their appreciation of the relevance of quantitative reasoning to us all” (“[Quantitative Reasoning Connection](#)”). My goal was to introduce students to ‘big data’ approaches to literary analysis that looks at a work, corpus of works, or database of works to identify patterns using computers and computational methods. Rather than looking at the specific language an author uses in a unique passage, distant reading asks students to consider larger patterns and conclusions from a broader spectrum of works. Ultimately, I want students to challenge their assumptions about how and why we do literary analysis. We often take our disciplinary practices at face value and don’t think too deeply about what they contribute or lack/ignore; they are simply how we do things. The conclusion to this assignment asks students to think about the practices of close reading and distant reading and to evaluate both practices in terms of how we approach, read, and understand literature. What do we gain from each approach? What do we lose or ignore? How might technology help us analyze literature in ways previously prohibitive? What do these approaches lack? Ideally, students will have a different, data-informed understand of the Sherlock Holmes canon, the texts that I used for my assignment, and will be able to think about their own practices more self-consciously in the future.

Supporting materials:

Course description, assignment with learning objectives, and teaching notes

Ann McClellan**Course Info: EN1600 Studies in English**

Catalog Description: This course is designed to introduce students to the major fields of study and debates in English studies today including, but not limited to, literary studies, film studies, education, writing, critical theory, aesthetics, cultural studies, and canon formation. The course will provide a brief introduction to some of the most recent developments in literary theory with a special emphasis on post-structuralist theories since the 1960s. We will evaluate how useful those tools are in understanding, interpreting, and appreciating literature, and we will analyze the ways in which these debates have influence both how and what we read in the 21st century college classroom. Students will be asked to internalize and incorporate these new skills and critical paradigms in class discussions and their own writing.

Course Learning objectives:

- a. Understand the social and historical context of English Studies
- a. Understand the options within the PSU English major and minor
- a. Understand literary periods, representative writers, and genres
- a. Understand literary terms, concepts, and theories
- a. Understand different ways to read and interpret literature
- a. Understand how to incorporate and document textual evidence and research

This course fulfills the Quantitative Reasoning Connection in the Disciplines (QRCO) requirement:

Mathematics finds application in all fields of scholarship. All disciplines make use of quantitative reasoning in some way and to some extent. Students take a three-/four-credit Quantitative Reasoning (Q) course specified as required for their major. This course may be taught within the major discipline or not. It might teach quantitative techniques used as primary or secondary tools within the discipline or might be a course in which students of less quantitative disciplines come to deepen their appreciation of the relevance of quantitative reasoning to us all.

Distant Reading Assignment: *The Adventures of Sherlock Holmes* (30 pts.)

Assignment Learning Objectives:

- understand the basic conventions of close reading and distant reading,
- learn about emerging data collection/quantitative techniques and research in English studies, and
- compare and contrast benefits and drawbacks (or explain methodological approaches) of close reading and distant reading

Context: We've spent the past few weeks in class learning about and practicing the basic tenets of close reading: identifying patterns within and across texts; looking at specific textual passages to identify themes/patterns, symbolism, connotation, etc.; and developing larger arguments, interpretations, and conclusions about specific passages from a text. We're now going to learn about another way to interpret texts, one that focuses more on the '30,000 foot' level, called Distant Reading.

You've done some reading on Distant Reading in class so far (Note: as this is a first-year, undergrad course, I want to keep these readings approachable so maybe this one from the [New York Times](#) or a quick definition guide from [UCLA](#)). For this assignment, we will look at an entire collection of the Sherlock Holmes stories instead of just one story at a time to identify and interpret patterns across an entire body of work. Like your close reading assignments, the distant reading assignment focuses on **pattern recognition** followed by **analysis** and **interpretation**.

Assignment:

Data:

Part A: COMPLETED. You already have the following data from our previous class discussions and assignments:

- 'Print' versions of the stories we read them for close reading discussion (could have them take picture of their 'print' annotations to see how they compare to digital models)
- Your own close reading analyses of the individual stories

Tools:

We will be using the online tool [Voyant](#) for this portion of the project. Voyant is a web-based tool that allows us to input and visualize large sets of text (aka, a corpus) in several different formats.

Part B:

- Upload the text file versions of the *Adventures of Sherlock Holmes*, which comprises the first 12 short stories Doyle wrote about the detective (<https://sherlock-holm.es/stories/plain-text/adv.txt>).
- Click "Reveal" to get results.

Experiment with at least TWO of the most provocative/ interesting/revealing tools and fully explore their capabilities. Keep in mind: you may not find anything particularly deep or revolutionary! But then again, you may. ☺ We're just experimenting here.

Saving:

Make sure to save/export your corpus/findings by clicking on the 'Export' (diskette) icon in the blue bar at the top, or you can export a link for an individual tool by clicking on the "Export" icon in one of the tool panes.

After **saving/exporting** your corpus findings, upload the results to the class blog and **write** a 700+ word post about the tools you used, the patterns you found, and the interpretations you came up with. You might want to focus on word frequency, pattern recognition, theme recognition, etc.

Pro-Tip: Think about what distant brings you that close reading does not (and vice versa). What can you see/learn/conclude from your distant reading that perhaps you did not think about in your original close readings? What does a digital humanities approach add to our more familiar individual text-based analyses of literature?

Evaluation/Grading:

This assignment is worth 30 pts (the same as a single close reading assignment).¹ As an experimental class project, you are not being graded on what you find about *The Adventures of Sherlock Holmes*. Instead, you'll be graded on (1) whether you accomplish all the parts of the assignment (pass / fail), (2) how engaged you are with the work, and (3) how well you apply the method of pattern recognition / interpretation we've been embracing throughout the semester.²

¹ For perspective, there are 1000 pts. total in the class as a whole, so this counts as 3% of their final grade.

² Portions of this assignment have been adapted and borrowed from [Brian Croxall's "Distant Reading Duffy" assignment](#).

Before students embark on the assignment/On day we review assignment in class:**In- Class Modeling:**

- a) Read through the "Getting Started" introduction on the Voyant website to familiarize ourselves with the tool, what it does, and how it works.
- b) Select the "Shakespeare's Plays" corpus as our starter.
- c) The various tools in the interface are designed to interact with one another. For instance, if you click on a word in Cirrus, you'll see the Trends tool update with information about the selected work. Click on "KING" in Cirrus and see how the Trends tool updates. What changed? At the bottom of the Trends tool, click "Reset" to set the graph back to the original findings.
- d) Similarly, if you click on a node in the Trends tool the Contexts tool should update as well. For example, "Sir" seems to occur a lot in Shakespeare. Click on the term "Sir" in the Trends graph and observe how the Contexts section (below) changes in response.
- e) Additional tools are readily accessible by clicking the tabs in each tool pane. For instance, beside the Cirrus header label is the Corpus Terms label, clicking on the tab will switch the tool. Tools readily available through the tabs are Corpus Terms, Links, Collocates, Documents, Phrases, and Bubblelines.
- f) **Stopwords:** Voyant pre-programs its tool to ignore/exclude many commonly used words that don't really add much to our analysis (the, a, an, and, or, etc.). You can add additional stopwords to the default list by clicking on the "Options" button in Cirrus in the top left. The pop-up window should show "Auto-detect." Next to that, click on the "Edit" button. This will bring you automatically to the bottom of the auto-detect list. You can add any words, followed by enter, to the list here. Let's try excluding "like" and "shall," both of which come up often in our initial word cloud. Click "Save" and check out how the word cloud and your other graphs have changed.
- g) **Bookmarking/Saving/Exporting:** Make sure to save/export your corpus/findings by clicking on the "Export" (diskette) icon in the blue bar at the top, or you can export a link for an individual tool by clicking on the "Export" icon in one of the tool panes.

Students can then play around with the tool in class and choose a song lyric, poem, or other text of their choice to upload into Voyant to see how this works and what they find.