CHAPTER 2

Distant Reading as Library Pedagogy:
Lessons for the Literary Studies Classroom

Amy Barlow

Introduction

The review of literature is a common feature of academic research and writing across the disciplines. In the humanities, it takes many forms, including but not limited to narrative reviews, bibliographic essays, historiographic essays, and reception history. I have observed through my experiences as a teaching librarian that these kinds of projects can leave undergraduate students feeling overwhelmed. My purpose in this chapter is to share a cooperative classroom approach, based on applications of the distant reading method, for engaging students in the difficult work of surveying the literature. I present conceptual background as well as practical examples of how to perform distant reading with bibliographic information provided by JSTOR and the *MLA International Bibliography*. The included lessons were created to help students see how their own ideas fit into the larger picture of scholarship on a topic. They may also help students identify perspectives that are absent from the literature. Equally important, distant reading can foster critical thinking about information resources, surfacing questions about the production and limits of bibliographic tools and underscoring the need for a plurality of resources

CC BY NC SA

during the research process. Finally, I argue that distant reading promotes learning about academic research in ways that are both enjoyable and supportive of pedagogical goals.

Critical Reading Connection

Distant reading is a philosophy and method originated by literary critic Franco Moretti and practiced within the digital humanities community for at least two decades.[1] It is a form of critical reading that can be enacted to decipher themes, theoretical approaches, or social and critical contexts over an extensive amount of material for various academic purposes.[2] Unlike close reading, which produces deep knowledge of a single text, distant reading relies on computational tools to gather and process information from a large number of texts. The reader then analyzes a summary of information (e.g., a graph or index) to make inferences and create meaning. In 2012, experiments conducted by information scientist David Mimno at Cornell University pointed to uses of distant reading for academic work. His analysis of a large collection of digitized classics journals demonstrated that distant reading can reveal patterns and anomalies in a complex corpus of secondary literature. While Mimno acknowledged that subject experts would find little to surprise them in an automated analysis of scholarship, he saw its advantages for inexperienced researchers, explaining that distant reading can be deployed to introduce a field of study and to root out texts that merit close reading.[3]

It is clear that students stand to benefit from distant reading at the undergraduate level. However, as Michael S. Seadle noted in one of the few discussions of distant reading in the LIS literature, "The initial learning curve for distant reading is non-trivial."[4] The available tools for extracting data from texts are scattered, unstandardized, and each developed for a specific type of analysis. Some demand computer programming abilities and many require a significant investment of time for customization and testing.[5] As teaching librarians, how might we introduce students to distant reading without imposing a high level of prerequisite skills or dedicating too much class time to training? One strategy is to lower the skills threshold through the use of a tool with which we are uniquely qualified to teach: the bibliographic database.

Johanna Drucker's *Introduction to Digital Humanities* defines distant reading as "the idea of processing content in (subjects, themes, persons, places, etc.) or information about (publication date, place, author, title) a large number of textual items without engaging in the reading of the actual text."[6] Because bibliographic databases contain an abundance of information about texts, there are researchers who turn to them for distant reading.[7] The authors of "'Spare your arithmetic, never count the turns': A Statistical Analysis of Writing about Shakespeare, 1960–2010" collected retrospective metadata, such as document type and names of plays, from records indexed by the World Shakespeare Bibliography Online. Their analysis showed quantitative evidence of historical patterns in Shakespeare scholarship, providing new directions for inquiry, such as opportunities in global Shakespeare studies. Like many practitioners of distant reading, the authors were careful to note that the method can lead to some unsurprising results: their data indicated that *Hamlet* was a popular critical subject.[8] In a different

study, Augsburg University professor Dallas Liddle gathered metadata from Cengage's *Times Digital Archive* in an attempt to draw correlations between PDF file size and typographical innovations in a large collection of periodicals. Although Liddle characterized his experiment as "barely above the back-of-the-envelope level," his purpose was to draw attention to the potential of bibliographic metadata for distant readings of Victorian text culture.[9]

Teaching Strategies

How can the kinds of distant reading experiments described in the preceding section be scaled for application in the classroom? Below you will find generalized instructions and two lesson plans that use a cooperative approach for performing the distant reading method with undergraduate students. In both cases, previous exposure to the concept of distant reading is beneficial, but not essential. In Lesson 1, students in a first-year seminar work in groups to explore and categorize bibliographic information culled from JSTOR. In Lesson 2, students in an English seminar work collectively to gather, visualize, and analyze bibliographic information from the *MLA International Bibliography*. Each lesson requires access to bibliographic tools and between 90 to 110 minutes of class time. Although both examples pertain to English studies, the techniques can be adapted to other disciplines by following these general steps:

1. Select a bibliographic database.
2. Set up search parameters.
3. Gather data.
4. Plot data.
5. Analyze results.
6. Reflect.

Lesson 1: Introductory Distant Reading

Students in an honors program first-year seminar, taught by English professor Maureen Reddy at Rhode Island College, were expected to gain an understanding of the critical conversation that surrounds Bram Stoker's *Dracula*. Professor Reddy organized students into work groups and asked them to find critical articles that might interest them and the entire class. Later in the semester, each group presented ideas from one article during an oral report. Professor Reddy requested that I meet with the class to teach students how to search for criticism. To that end, I developed a lesson that connected distant reading with the concept of Searching as Strategic Exploration from the ACRL *Framework for Information Literacy for Higher Education*. My goals were for students to "match information needs and search strategies to appropriate search tools" and to "recognize the value of browsing and other serendipitous methods of information gathering" through a simple process of collecting and categorizing bibliographic metadata discovered in JSTOR.[10] The lesson anticipated that students in the class would know more than me about *Dracula* as a cultural text because it was the entire focus of their course.

Preparation

This lesson requires little prep time. First, you will want to know a little bit about the subject matter. Second, construct test searches in JSTOR with the goal of selecting keywords and parameters that produce multiple pages of relevant results. Third, adapt the lesson below to create a worksheet in a collaborative web-based tool such as Google Docs.

Learning Outcomes

By the end of this lesson, students will be able to do the following at an introductory level:

- Identify JSTOR as a tool for finding academic publications.
- Construct an effective search for criticism of *Dracula* in JSTOR.
- Select and categorize topical information about a large number of publications.
- Select appropriate publications for close reading.

Lesson Plan

1 - Introduce the activity by briefly explaining why researchers use bibliographic databases. Describe why it's advantageous to spend time exploring a large number of results, looking for patterns, rather than just focusing on the top few articles in a results list. Provide an overview of the in-class activity and articulate how it will help students to meet the requirements of their assignment.

2 - Collect data. Divide students into their work groups. Show them where to find the JSTOR database on the library website. Show the class how to construct a JSTOR search for *Dracula* AND *Bram Stoker* with limits on language (English), publication date (1999–2020), and discipline (Film Studies and Language & Literature). Assign each of the workgroups one of four pages of results to review. Provide the class with a link to a shared Google Doc with open editing privileges. Ask students to read the article titles, journal titles, and topics/subject terms that appear on their assigned results page in JSTOR.[11] As they review the metadata, ask each student to use the Google Doc to list one or two subject and/or topic examples from the scholarship under each of the following categories (repetition of topics is fine):

- Expected Topics
- Surprising Topics
- Unfamiliar Topics

3 - Discuss. Look at the document together as a class and ask students to explain a few of the topics listed under each category. Point out repetition of subjects and encourage students to make inferences about popular and outlier interpretations of *Dracula*. Ask students about the unexpected topics and use the list to select a full-text article to display in front of the class. Evaluate the full-text publication against the requirements of the assignment (e.g., page length, publication date, etc.). Define unfamiliar topics with the assistance of the faculty member who teaches the course. If a student sees a topic on the list that they wish to know more about, ask them to connect with the student who put the information into the list so that they can work together to obtain the full-text article.

4 - Reflect. Ask students: How might you use this list of topics to further your search for *Dracula* criticism? They can put their thoughts into the Google Doc.

Lesson 2: Intermediate Distant Reading

Students in *Literature by Women*, an English seminar taught by Rhode Island College professor Barbara Schapiro, came to the library for a class on how to find literary criticism for a final paper. Because students were expected to report on some of the important qualities in the scholarship, such as main ideas and critical approaches, I thought that a distant reading activity would prove useful. Working in groups through a series of prompts, students collected quantitative and qualitative metadata from a group of bibliographic records indexed by EBSCOhost's *MLA International Bibliography* (MLAIB). Each group analyzed a subset of the data and reported their findings to the class. I charted the results on the board as a way to visualize a record of scholarship on a required text, Jane Austen's *Sense and Sensibility*. Discussion followed with key insights from Professor Schapiro.

Preparation

This lesson demands a fair amount of preparation. First, in order to interpret results with the class, you will want to be knowledgeable about the history, scope, and structure of the featured bibliographic index. Background information is sometimes available on the publisher's website, or you could seek out database performance reviews.[12] Second, it is helpful to study a few distant reading experiments that use metadata provided by library databases, such as those referenced earlier in this chapter. Third, discuss the distant reading method with the professor of the course. They may be unfamiliar with the practice and eager to ask questions. If the professor is knowledgeable about distant reading, they may be receptive to preparing students for the lesson by introducing the concept prior to the library session. Fourth, construct test searches in the MLAIB to see how many publications are indexed and over what time period. Distant reading is useful for summarizing dozens, hundreds, or thousands of texts—not so much a handful of texts. Fifth, prepare a worksheet for students.

Learning Outcomes

By the end of this lesson, students will be able to do the following:

- Define distant reading.
- Construct an effective search for criticism of a literary work in the MLAIB.
- Summarize and analyze bibliographic information.
- Relate patterns in the indexed literature to social and academic movements.

Lesson Plan

1 - Introduce the activity by explaining to students that they will enact distant reading with the *MLA International Bibliography*. Spell out the purpose of distant reading in

relation to the coursework. In the *Literature by Women* seminar, we were distant reading the MLAIB to reveal critical movements in the study of *Sense and Sensibility* for a major writing assignment. Provide an overview of the activity and assure students that you will show them how to set up the search parameters.

2 - Collect data. Divide students into decade groups (i.e., 1970–1979, 1980–1989, etc.). Go to the MLAIB database and show the class how to construct a search for *Sense and Sensibility* as a Primary Work with limits on Publication Date. Provide students with a worksheet and ask them to read the bibliographic records—not the full-text articles—in their decade and to gather the following information from the metadata:

- total number of publications in your decade
- types of publications (e.g., books, book chapters, journal articles, dissertations, or other)
- languages
- popular or repeated subjects
- subjects that indicate a critical approach or theory
- an unfamiliar or unexpected subject
- a subject that interests you personally

3 - Visualize data. Use a whiteboard or digital tool to create a graph titled "The Study of *Sense and Sensibility*". Label the horizontal axis "date of publication" and label the vertical axis "total publications". Ask each group to share the total number of publications indexed during their decade. Plot the data on the graph to form a trend line. Beginning with the earliest decade group, ask students to share information gleaned from the metadata. Add selected information to the graph above the corresponding data point for the decade. (See figure 2.1.)

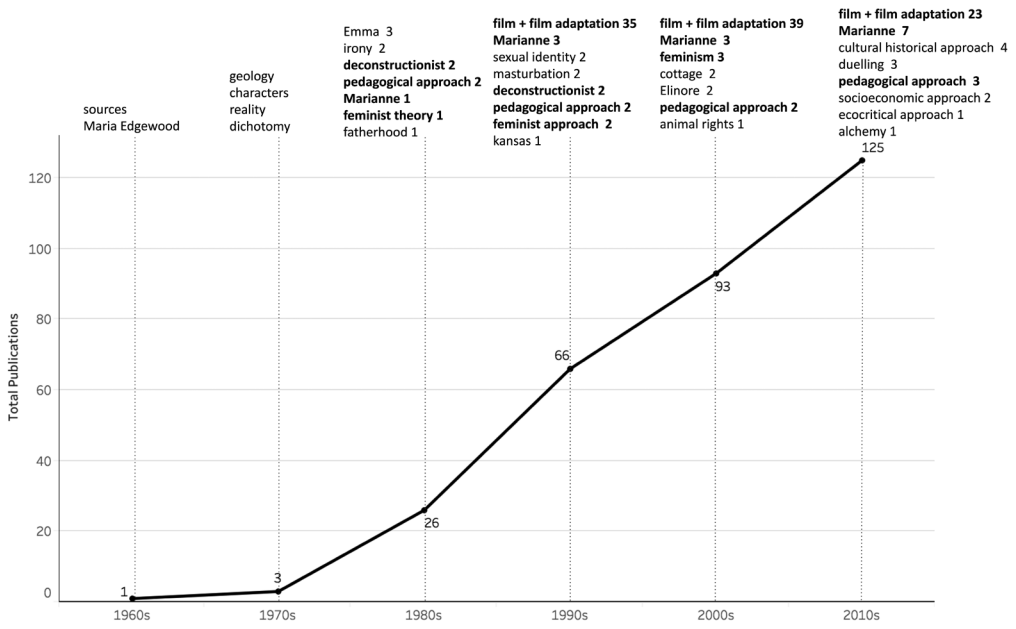

Figure 2.1

The study of *Sense and Sensibility*, 1960–2020, with metadata from EBSCO's *MLA International Bibliography*. Bold subject headings appear across several decades.

4 - Discuss. The subject expertise of the faculty teaching the course is crucial for interpreting results. Look at the graph together as a class and begin by posing a few questions. What, if anything, did we learn about trends in writing about *Sense and Sensibility*? When do we see a dramatic increase in scholarship? When was the first entire book about *Sense and Sensibility* published? What was the impact of film adaptations? What subjects and critical approaches have been popular with scholars? What critical approaches are absent from the scholarship? Which characters have been studied and which have received less attention? What unexpected or unusual subjects did you encounter? How might all of this information be useful to you while doing research for your final paper?

5 - Reflect. If time permits, it is worthwhile to hear students' thoughts on distant reading. What did you learn about *Sense and Sensibility* that was new to you? How does distant reading compare with reviewing and selecting full-text publications during class? What are the benefits and flaws of this exercise? Do you think that the distant reading technique will be helpful to you in future research papers?

Discussion

In my experience, lessons based on distant reading are enjoyable and useful. For instance, the students in Professor Reddy's first-year seminar were delighted to display their budding knowledge of *Dracula* as a cultural text by calling out "expected" critical themes, such as *Gothic*, *science*, and *sexuality*. They highlighted the ingeniousness of literary interpretation when they noticed such "surprising" subjects as *clocks*, *circumcision*, and *veganism*. Their choice of "unfamiliar" topics, such as *deconstructionism*, *Bergsonism*, and *narrative authority* revealed gaps in their knowledge. Professor Reddy joined the Google Doc to participate in the exercise. She explained some of the patterns in the criticism, defined unfamiliar disciplinary terms, and provided guidance as students selected topics from the list to identify appropriate articles for close reading. In response to the end-of-class question about how distant reading can help researchers, one group of students wrote that the activity exposed them to "different" and "abstract" topics that they would not have known to look for on their own. They added, "The more you search, the more interested you become."[13]

In upper-level courses, where students have more subject knowledge and experience with library databases, distant reading can support critical pedagogy by developing students' awareness of the power relations built into bibliographic classification systems and database tools. As teaching librarians, we can design lessons that facilitate learning about the racial, ethnic, and gender biases built into information resources.[14] Our classroom discussions can surface those disparities so that students are motivated to use a plurality of tools during the research process. Take, for example, the visualization of writing about *Sense and Sensibility*. If publications on the topic are scarce before the 1960s, you might explain that between 1922 and 1957, the MLAIB generally indexed American scholarship only. This provides an opportunity to suggest additional resources, such as the library catalog, JSTOR, or the *Annual Bibliography of English Language and Literature*, for finding earlier criticism of the literary work. Or you might ask students to compare

the availability of descriptive subjects before and after the 1980s, calling attention to the invention of electronic indexing, and asking students how they might adjust their search constructions to account for the differences. We can raise questions about content: What types of publications and subject matter are included? What types are excluded? What criteria are used to make determinations? Where do we find materials that fall outside of the criteria?

Distant reading serves various purposes and provides a fresh take on learning about important aspects of academic research, but in the classroom, I always keep Dallas Liddle's perspective in mind—these are "back-of-the-envelope" experiments.[15] The exercises described in this chapter use retrospective bibliographic information instead of data-mining the actual text. There is room for error in the search term construction, data collection, and analysis. Even taking into account a wide margin of error, however, distant reading generally gives undergraduate students new insights into the nature and scope of writing on a subject. It's an unexpected critical reading approach that encourages active participation in the classroom, enabling students to work cooperatively to construct their own knowledge of disciplinary literature and to relate their ideas to past and current thinking on a topic. Although most students will continue to visit databases for full-text content, those who have practiced distant reading may be tempted into *not* immediately going for the articles in favor of a critical reading method that brings a bigger picture of the literature into focus and inspires questions about how our library resources are produced and mediated.[16]

Conclusion

Despite attention from scholars in other fields of study, the LIS literature offers limited discussion of how distant reading benefits pedagogy or the advantages of utilizing bibliographic tools in the process. This chapter introduced distant reading with information provided by bibliographic databases as an approach for drawing out the essential qualities of scholarship on a topic and for satisfying other pedagogical goals. While the techniques that I described in this chapter were developed for English studies, it would be possible to duplicate them in other subject areas using different databases. As librarians, we have a terrific opportunity to show students and faculty colleagues that our existing bibliographic resources can serve as tools for distant reading and perhaps even other digital humanities methods.

Notes

1. Franco Moretti, "Conjectures on World Literature," *New Left Review* (2000): 54–68. For historical perspective on distant reading, see Ted Underwood, "A Genealogy of Distant Reading," *DHQ: Digital Humanities Quarterly* 11, no. 2 (2017).
2. For an overview of critical reading purposes, see Karen Manarin et al., *Critical Reading in Higher Education: Academic Goals and Social Engagement* (Bloomington: Indiana University Press, 2015): 4–8.
3. David Mimno, "Computational Historiography: Data Mining in a Century of Classics Journals," *Journal on Computing and Cultural Heritage* 5, no. 1 (2012): 1–19.
4. Michael S. Seadle, "Managing and Mining Historical Research Data," *Library Hi Tech* (2016): 177.

5. Seadle, "Managing and Mining."
6. Johanna Drucker, David Kim, Iman Salehian, and Anthony Bushong, *Introduction to Digital Humanities: Concepts, Methods, and Tutorials for Students and Instructors* (University of California Los Angeles, 2014), https://searchworks.stanford.edu/view/11649226.
7. The studies referenced in this chapter utilize extant metadata from proprietary library databases. For an example of distant reading with open access bibliographic records, see J. Stephen Downie, Sayan Bhattacharyya, Francesca Giannetti, Eleanor Dickson Koehl, and Peter Organisciak, "The HathiTrust Digital Library's Potential for Musicology Research," *International Journal on Digital Libraries* (2020): 1–16.
8. Laura Estill, Dominic Klyve, and Kate Bridal, "'Spare Your Arithmetic, Never Count the Turns': A Statistical Analysis of Writing about Shakespeare, 1960–2010," *Shakespeare Quarterly* 66, no. 1 (2015): 1–28.
9. Dallas Liddle, "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' *The Times* using *The Times Digital Archive*," *Journal of Victorian Culture* 17, no. 2 (2012): 230.
10. *Framework for Information Literacy for Higher Education*, American Library Association, February 9, 2015, http://www.ala.org/acrl/standards/ilframework.
11. At the time of this publication, JSTOR's topics and underlying thesaurus had disappeared from the database. On January 8, 2021, a Twitter communication from JSTOR indicated that the company is currently working on the thesaurus. JSTOR expects to reintroduce this feature after improvements.
12. For an explanation of the history and scope of the MLAIB, see Vince Graziano, "Retrieval Performance and Indexing Differences in ABELL and MLAIB," *Journal of Electronic Resources Librarianship* 24, no. 4 (2012): 268–87.
13. Group 3, Worksheet for the "Exploring *Dracula* Criticism" Library Workshop, Honors 100, Rhode Island College, October 30, 2019.
14. For background, see Richard Jean So and Edwin Roland, "Race and Distant Reading," *PMLA* 135, no. 1 (2020): 59–73; Kim Gallon, "Making a Case for the Black Digital Humanities," *Debates in the Digital Humanities*, ed. Matthew K. Gold and Lauren F. Klein (University of Minnesota, 2016): 42–49.
15. Liddle, "Reflections on 20,000 Victorian Newspapers," 8.
16. To quote Moretti, 57: "We know how to read texts, now let's learn how *not* to read them."

Bibliography

American Library Association. *Framework for Information Literacy for Higher Education*. February 9, 2015. http://www.ala.org/acrl/standards/ilframework

Downie, J. Stephen, Sayan Bhattacharyya, Francesca Giannetti, Eleanor Dickson Koehl, and Peter Organisciak. "The HathiTrust Digital Library's Potential for Musicology Research." *International Journal on Digital Libraries* (2020): 1–16.

Drucker, Johanna, David Kim, Iman Salehian, and Anthony Bushong. *Introduction to Digital Humanities: Concepts, Methods, and Tutorials for Students and Instructors*. University of California Los Angeles, 2014.

Estill, Laura, Dominic Klyve, and Kate Bridal. "'Spare Your Arithmetic, Never Count the Turns': A Statistical Analysis of Writing about Shakespeare, 1960–2010." *Shakespeare Quarterly* 66, no. 1 (2015): 1–28.

Gallon, Kim. "Making a Case for the Black Digital Humanities." *Debates in the Digital Humanities*, edited by Matthew K. Gold and Lauren F. Klein, 42–49. University of Minnesota, 2016.

Graziano, Vince. "Retrieval Performance and Indexing Differences in ABELL and MLAIB." *Journal of Electronic Resources Librarianship* 24, no. 4 (2012): 268–87.

Liddle, Dallas. "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' *The Times* using *The Times Digital Archive*." *Journal of Victorian Culture* 17, no. 2 (2012): 230–37.

Manarin, Karen, Miriam Carey, Melanie Rathburn, and Glen Ryland. *Critical Reading in Higher Education: Academic Goals and Social Engagement*. Bloomington: Indiana University Press, 2015.

Mimno, David. "Computational Historiography: Data Mining in a Century of Classics Journals." *Journal on Computing and Cultural Heritage* 5, no. 1 (2012): 1–19.

Moretti, Franco. "Conjectures on World Literature." *New Left Review* (2000): 54–68.

Seadle, Michael S. "Managing and Mining Historical Research Data." *Library Hi Tech* (2016): 172–79. https://doi.org/10.1108/LHT-09-2015-0086.

So, Richard Jean, and Edwin Roland. "Race and Distant Reading." *PMLA* 135, no. 1 (2020): 59–73.
Underwood, Ted. "A Genealogy of Distant Reading." *DHQ: Digital Humanities Quarterly* 11, no. 2 (2017).