Plan of Study

In literary study, scholars increasingly embrace the methods of "distant reading" to study hundreds or thousands of texts using computational tools. As an alternative to "close reading" (the practice of meticulously reading and researching the language of small selections of individual texts), distant reading promised an exciting opportunity to expand the borders of which literature would merit serious attention. However, in the nearly twenty years since Franco Moretti coined the term "distant reading" in his "Conjectures on World Literature," the introduction of computation has not changed the preference for a limited, selective canon. A recent article by Kathleen Bode exemplifies a current shift in digital humanities research. Bode critiques pioneers of distant reading and literary macroanalysis Franco Moretti and Matthew Jockers, arguing that they too often disregard the provenance of the thousands of texts they use as proof of their arguments. As an alternative, Bode calls for an integration of computational research with the textual scholarship of fields like bibliography and book historym which she terms a "data-rich literary history." My dissertation answers this call by systematically researching all texts printed in England between the years 1789 to 1799, to redefine our understanding of the definitive works of this decade.

Despite the crucial importance of corpus-building (that is, the accumulation of a large body of text files for analysis) to the interpretation of "distant reading" research, it is often extremely difficult to know what is in a corpus. Even large institutional resources used by many scholars provide little context for their choices of what to include or exclude. These hidden choices are particularly problematic when historical selection factors might have led to the creation of databases which re-create social inequalities. I focus specifically on writing printed in England between 1789 and 1799, to explore how works from this eleven-year "decade" have been selected as important, literary, or popular. For this period, the English Short Title Catalogue provides basic bibliographic data for nearly 52,000 titles, but the Eighteenth Century Collections Online Text Creation Partnership corpus of XML-encoded full texts — the source for most distant reading research on eighteenth century literature — includes fewer than 500 titles. This difference raises the question: why were the other 51,500 titles not considered worth the investment of scholarly effort? And with particular urgency: do the most invested-in resources underrepresent women?

My experiments examine six major databases to answer these questions: The English Short Title Catalogue (ESTC), Eighteenth Century Collections Online (ECCO), the Eighteenth Century Collections Online Text Creation Partnership (ECCO-TCP), Google Books, Project Gutenberg, and HathiTrust. For each database, I download their holdings identified as printed in England 1789-1799. I calculate how many works are attributed to male, female, or unknown authors. I then examine the contents of each database more closely, to compare the inclusion of broad categories of writing like poetry, drama, prose fiction, sermons, non-fiction, and ephemera. Identifying these categories of writing within each corpus reveals a predictable preference for "literary" forms such as novels and poetry (rather than non-fiction or sermons) in the smaller databases. This preference for particular kinds of writing might explain changes in gender representation. If novels are mostly written by women, for example, a corpus can underrepresent women by underrepresenting novels. Or it could include a representative number of novels, but disproportionately include novels by men. My investigation allows me to identify the patterns of selection.

I have completed my data collection, begun writing the necessary code to carry out my experiments, and written my first chapter. The Ontario Graduate Scholarship would fund the sixth year of my PhD, during which I will complete code development and write two chapters.

Bibliography

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; Or, Towards a New Object for Data-Rich Literary History." *Modern Language Quarterly* 78.1 (December 2017): 77-106.

—-. "Why You Can't Model Away Bias." Preprint: *Modern Language Quarterly*, volume 80, number 3, forthcoming 2020.

Brewer, David. "Counting, Resonance, and Form, A Speculative Manifesto." *Eighteenth-Century Fiction*, volume 24, issue 2, 2011, pp. 161-170.

Clery, E.J. *The Rise of Supernatural Fiction 1762-1800*. Cambridge UP, 1995.

Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. Harvard UP, 2014.

Farmer, Alan, and Zachary Lesser. "Structures of Popularity in the Early Modern Book Trade." *Shakespeare Quarterly* 56.2 (2005): 206–13.

Felski, Rita. *The Limits of Critique*. U Chicago, 2015.

Fischer-Starcke, Bettina. *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*. Continuum, 2010

Gamer, Michael. *Romanticism and the Gothic: Genre, Reception, and Canon Formation*. Cambridge UP, 2000.

Garside, Peter et al., eds. *The English novel, 1770-1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*. Oxford UP, 2000.

Gavin, Michael. "Historical Text Networks: The Sociology of Early English Criticism." *Eighteenth-Century Studies*, volume 50, number 1, 2016, pp. 53-80.

Gilmartin, Kevin. *Print politics: The Press and Radical Opposition in Early Nineteenth-Century England*. Cambridge UP, 1996.

Guillory, John. *Cultural Capital: The Problem of Literary Canon Formation*. U Chicago, 1993.

Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. U Illinois, 2013.

Klancher, Jon P. *The Making of English Reading Audiences, 1790-1832*. U Wisconsin, 1987.

Lahti, Leo, Niko Ilomäki, and Mikko Tolonen. "A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800." *LIBER Quarterly*, volume 25, number 2, 2015, pp. 87–31.

Love, Heather. "Close but Not Deep: Literary Ethics and the Descriptive Turn." *New Literary History*, volume 41, issue 2, 2010, pp. 371–391.

Marcus, Sharon, Heather Love, and Stephen Best. "Building a Better Description." *Representations*, volume 135, number 1, Summer 2016, pp. 2-21.

McCarty, Willard. "Knowing: Modeling in Literary Studies." *A Companion to Digital Literary Studies*, ed. Susan Schreibman and Ray Siemens. Blackwell, 2008.

Mee, John. *Print, Publicity, and Popular Radicalism in the 1790s*. Cambridge UP, 2016

Moretti, Franco. "Conjectures on World Literature." *New Left Review*, volume 1, issue 1, 2000, pp. 54 - 67.

—-. *Distant Reading*. Verso, 2013.

Samuels, Lisa and Jerome McGann. "Deformance and Interpretation." *New Literary History* volume 30, number 1, Winter 1999, pp. 25-56.

Sedgwick, Eve. "Paranoid Reading and Reparative Reading, or, You're So Paranoid, You Probably Think This Essay Is About You." *Touching Feeling*. Duke UP, 2003.

St. Clair, William. *The Reading Nation in the Romantic Period*. Cambridge UP, 2004.

Worrall, David. *Radical Culture: Discourse, Resistance and Surveillance, 1790-1820*. Harwester Wheatsheaf, 1992.