

## The Doxa of Reading

---

ANDREW GOLDSTONE

Reading *Graphs, Maps, Trees* as a late-stage graduate student in 2008 was invigorating.<sup>1</sup> Here was an approach to literary history free from the pieties of “close reading,” committed to empiricism, seeking to fulfill the promise of the sociology of literature with its “materialist conception of form” (92). And, at the time, it seemed entirely natural that the way to follow the path laid out by Franco Moretti in *Graphs* and in the essays he had published over the previous decade was to go to my computer, polish my rusty programming skills, and start making graphs. Yet reconsidering *Distant Reading* now, it is striking how non-digital the book is. In fact, the term “distant reading” has undergone a rapid semantic transformation. In “Conjectures on World Literature,” originally published in 2000, Moretti introduces the phrase to describe “a patchwork of other people’s research, *without a single direct textual reading*” (*Distant* 48). Today, however, “distant reading” typically refers to computational studies of text. Introducing a 2016 cluster of essays on “Text Analysis at Scale,” Matthew Gold and Lauren Klein speak of “using digital tools to ‘read’ large swaths of text” as “‘distant reading’” (Introduction); in his contribution there, Ted Underwood embraces “distant reading” as a name for applying machine learning techniques to unstructured text (“Distant Reading”). Discussions of distant reading have become discussions of computation with text, even if no section of *Distant Reading* features the elaborate computations found in the Stanford Literary Lab pamphlets to which Moretti has contributed.

The semantic shift takes about ten years. At first—judging from the 190-odd results for the phrase in a JSTOR search in early 2017—invocations of the term are concentrated in debates about world literature (e.g., Muftić 487). But soon enough, computers enter the picture: in 2007, Ana Mitrić published

---

1. This paper has benefited from exchanges with Anne DeWitt and Ted Underwood.

a computer-aided study of Jane Austen's lexicon subtitled "A Distant Reading"; in 2008 the historian Daniel Cohen envisions studying "a very accurate and complete database of every single use of the Bible in the Victorian era" and terms this "distant reading" (481). Best and Marcus's 2009 "Surface Reading" uses the term (without citing Moretti) as an example of a field turning towards "computers, databases, and other forms of machine intelligence" (17). Then the mass-circulation press throws its weight behind the new meaning: a 2011 *New York Times Book Review* piece headlined "What Is Distant Reading?" discusses "hypothesis-testing, computational modeling, quantitative analysis" in the first two Literary Lab pamphlets (Schulz). Today, it is possible to refer to "distant reading" *in the context of a world-literature discussion* as self-evidently computational; in 2015, Simon During, describing an approach to world literature "which we associate with Franco Moretti," says it "bypasses criticism and evaluation, calling upon 'distant reading,' dependent on digitized data-bases" (35). The computers have won, symbolically speaking.

This trajectory is also Moretti's own, as implied by the reference in *Distant Reading's* last essay (dated 2011) to the importance of "digital databases and automated data retrieval" (212). The headnote to "Style, Inc." retraces his interests: "one day I realized that the study of morphological evolution had itself morphed into the analysis of quantitative data" (179). But the move from "data" to computers is yet a further leap. "Style, Inc." and "Network Theory, Plot Analysis" are quantitative analyses more or less by hand; "The Novel: History and Theory" (dated 2007) likewise leaves for the future "a quantitative stylistics of the digital archive" (165). Within *Distant Reading*, "distant reading" seems to refer to Moretti's effort, in most of the essays, to synthesize his own and other scholarly readers' interpretations of texts into hypotheses about literary-historical systems.

Though the work subsequently done under the "distant reading" rubric has been lively and varied, the changing meaning of the phrase has obscured rather than answered some of the most significant questions raised in *Distant Reading* (and not only there) about the methods and aims of literary study. Moretti's work joined a current of historicist and sociological challenges to what I call, after Pierre Bourdieu, the *doxa of reading*: the assumption that the primary activity of academic literary study is textual interpretation. Under this assumption, "reading" includes both the act of reading expertly and the production of expert readings of texts in articles and books. *Distant Reading* sometimes challenges the centrality of "reading" to the study of literary and social systems. But this challenge has been blunted in two ways: it has been misrecognized in terms

of a confrontation between “close” and “distant” techniques of “reading,” and it has been displaced from the research agenda by by developments in the practice of distant reading, which I will polemically summarize as *textualization*, driven by a return to national frameworks and the increasing predominance of the *text corpus* as analytical object. These developments, I will show, are also anticipated in the latest essays in *Distant Reading*, but they should not be regarded as the natural outcome of Moretti’s challenge to the doxa of reading. By reformulating the proposal for a major disciplinary realignment as a question of what can be learned from computational readings of monolingual text corpora, the practice of “distant reading” has yet to meet the strongest demands of *Distant Reading*.

#### BREAKING WITH THE DOXA

In *The Rules of Art*, Bourdieu defines doxa as what participants in the literary field take for granted, the assumptions by which “cultivated people are in culture as in the air they breathe” (185). For literary scholars, the doxa is consists in belief in the inherent value and meaning of the literary object, “rooted in the job and ethos of the professional commentator on texts” (194). Bourdieu argues that the literary field accommodates *both* orthodoxy and heterodoxy about this assumption as moves in the social game. Neither achieves a comprehensive understanding of the field itself: research risks being “imprisoned within the alternatives of an enchanted cult or a disabused denigration” (184). The tendency to characterize every method in literary study as a method of reading (surface, close, symptomatic, reparative . . .) indicates that this literary doxa persists today. Bourdieu argues that the “scientific objectification” (184) of the literary field requires not simply the negation of doxa but a “rupture” (185) that suspends the question of whether it is true or false.

At times, *Distant Reading* seems to carry out this rupture with doxa, especially in the two manifestos from 2000, “Conjectures on World Literature,” and “The Slaughterhouse of Literature.” Each takes a literary and social *system* as the object of analysis: the global novel or the field of *Strand* detective stories in the 1890s. Both essays eschew textual commentary, but they do not simply switch from close reading to some kind of anti-close reading. Instead, Moretti’s method aggregates interpretations of many texts. “Conjectures” may be considered an informal data analysis based on Moretti’s distillation of a large body of scholarship. “Distant reading” here *is not a kind of reading at all* but a search for empirical patterns validated through the consideration of many compara-

ble cases. Moretti, characteristically, speaks of a “*law of literary evolution*” (50). Similarly, in “Slaughterhouse,” all the stories are read, but none of them are quoted: instead, their use of clues is coded and categorized, and Moretti hypothesizes about the market role of these clues.

The data of “Slaughterhouse” correspond quite closely to what the social sciences call a content analysis. Content analysis formalizes the procedure of anatomizing the contents of texts and studying the aggregated results.<sup>2</sup> The reading of primary texts is *not* excluded: it remains essential, but only as a step towards the aggregate analysis, the search for pattern, law, and explanation. Moretti himself reflects on the source reading he practices, in these two essays, as a reading under constraint: in “Conjectures,” “a reading without freedom” that “merely *tests*” interpretations (53n19); in “Slaughterhouse,” a reading “in light of *something*” (87). *Graphs* calls it “a preference . . . for the explanation of general structures over the interpretation of individual texts” (91). By treating reading as a means rather than an end, Moretti challenges us to understand literary knowledge production in terms unrecognizable within the doxa of reading.

This shift—even more than the use of quantitative methods alongside qualitative ones—aligns Moretti with the social sciences, and with sociology in particular. “Sociological formalism has always been my interpretive method,” he says (59); the “common denominator” of *Graphs, Maps, Trees* is “a *materialist conception of form*” (92). Yet Moretti’s interest in evolutionary theory and the *Annales* as models means he only occasionally looks to scholarship analyzing literary systems sociologically. “Conjectures” and “Slaughterhouse” do cite Margaret Cohen for the important phrase “the great unread,” and “Conjectures” acknowledges Pascale Casanova; both these scholars in turn use Bourdieu’s framework, explicitly breaking with the norms of literary reading in order to analyze the field. One might equally compare “Conjectures” and “Slaughterhouse” to the work of Wendy Griswold and Janice Radway. Griswold’s 1981 study of the late-nineteenth-century American novel anticipates many features of those two essays; it analyzes thematic and formal features across a large sample of novels. Radway’s foundational *Reading the Romance* complements its ethnographic and book-historical evidence with a “Proppian” analysis of a sample of romances (120). Both Radway’s and Griswold’s argu-

---

2. Krippendorff’s *Content Analysis* is a standard handbook. The sociologist Tressie McMillan Cottom, in her critical discussion of Moretti, discusses a small-scale content analysis in order to suggest the difference between a sociological “distant reading” and the kind discussed in DH circles.

ments require aggregation, rather than single-text interpretation, to support their social-structural claims.

Yet comparing Moretti to these other scholars also helps to see why “distant reading” has not always registered as a challenge to the doxa of reading. Radway, especially, suspends doxa by taking *non-expert* reading seriously as an object of analysis. When Moretti makes a synthesizing claim about what readers do, he does not produce direct evidence: he attributes the triumph of the clue in detective stories to readers’ formal preferences, but this hypothesis is supported only by citations to an economic article on movie blockbusters and Walter Benjamin (69–70).<sup>3</sup>

Then there is the phrase “distant reading” itself. As an antonym of “close reading,” distant reading could also be taken as a heterodoxy rather than a rupture, in Bourdieu’s terms. Moretti strengthens the impression by calling it a “little pact with the devil” (48) and scorning close reading as “secularized theology, really (‘canon!’)” (67). Moretti’s characterization of New Haven-style close reading, though entirely accurate, positions him as an opponent of orthodox reading rather than someone who refuses reading as a master term. This anticipates a contemporary scholarly discourse in which, as Katherine Bode has argued, “close reading and distant reading are not opposites” (79), because they both tend to take their textual objects for granted. For Bode, this tendency stems from the neglect of bibliography and textual studies; this neglect, I suggest, is yet another consequence of the doxa of reading. It is due to a logic that operates across the whole field, not just within the domains where “distant reading” is discussed.<sup>4</sup> Participants in the field accommodate themselves more easily to an understanding of distant reading as a heterodox kind of reading, rather than as a rupture.

---

3. Later, in “The End of the Beginning,” Moretti speculates that cognitive science could supply the missing mechanism (142–43); this suggestion hints at a preference for a universalized reader, one who could be more easily identified with the expert critic. For an analysis of Victorian genre in which the data come from the diverse responses of actual readers, see DeWitt.

4. Meredith McGill points out that historicist literary criticism writ large, in “treating culture itself as a text,” “largely takes for granted the singularity and stability of literary texts” (20). My diagnosis of distant reading is quite similar to Bode’s; I lack space here to discuss her prescription for a “scholarly edition of a literary system” (102).

Distant reading becomes a kind of reading when its objects are seen as texts. Two further tendencies in the short history of “distant reading” abet this transformation: its progressive dissociation from comparative problems and its increasing concentration on text corpora. I will consider the former tendency only briefly. As “distant reading” diffused into English and other single-language disciplines, its typical object was shaped by the national specialization of most literary scholars: typical recent examples are Underwood and Sellers on British poetry (“*Longue Durée*”) and Wilkens on American novels (“Geographic Imagination”). It has become easy to forget that “Conjectures” proposed “distant reading” as a solution to two distinct problems: world literature *and* the “great unread” (45). World literature is a problem because it consists of a diverse system of literatures that *someone has to read* with sufficient interpretive acumen to discern what Moretti calls, in “More Conjectures,” “significant facts” (111). By contrast, the “great unread” is the problem of mass print in even one national system, consisting of what *cannot possibly be read*: “there are thirty thousand nineteenth-century British novels out there, forty, fifty, sixty thousand . . . no has read them, no one ever will” (45). The shared Britishness of all those nineteenth-century novels invites a different methodology than does the world-literary system as Moretti conceives it; the population of Victorian novels cannot make the same claims to attention—which is to say, to *worth* in the academic field—that the constituent literatures of the world can.

For “Slaughterhouse,” Moretti *did* read his sample of rivals to Conan Doyle. But even within *Distant Reading*, when Moretti returns to the single-nation framework with “Style, Inc.,” it is no longer even possible to read all seven thousand texts. Nor is there a scholarly literature about them to be aggregated; the titles themselves form the object of analysis. In that analysis, “mediocre conservative writers” are key agents in Moretti’s proposed explanation for a major “perceptual shift” in novels’ self-presentation (202). The widespread mediocrity of all those novels is also an *enabling condition* for the method; as Natalia Cecire has pointed out, the great unread starts to look like a mass of texts that are mostly not worth reading in the first place (305n10). It is difficult to imagine anyone making the same judgment about entire literatures, even “minor” ones, without referring to someone’s reading. The limitation to Great Britain permits us to feel that we already know enough about the literary history in question to supply the necessary context.

This assumption, by dispensing with *readings* as an intermediate step, per-

mits the return of *reading* in “Style, Inc.” by means of a change of object for literary interpretation: the *text corpus* of titles. Here, *Distant Reading* does anticipate the development of “distant reading.” Full-text corpora have become almost the exclusive object of analysis for work following on Moretti’s: of the thirteen Literary Lab pamphlets, nine are corpus studies (two thirds of these are about British novels); so are most other prominent efforts in quantitative literary history, including—to name only three examples—Underwood and Sellers, the bulk of Jockers’s *Macroanalysis*, and Long and So’s recent work, exemplified by their study of English haiku (“Literary Pattern Recognition”).

Though *Distant Reading* does not pursue the complex computations of later work on corpora, “Style, Inc.,” like “Network Theory, Plot Analysis,” anticipates important features of the current form of distant reading: the essay sustains its points through an interpretive discourse endowing the particulars of a single textualized object with generalizing force. In a characteristic move, Moretti discusses a transition from titles like *The Fakeer* and *The Vampyre* to *The Unfashionable Wife* and *The Discarded Daughter*:

Without adjectives, we are in a world of adventures; with adjectives, in a destabilized domesticity. The adjective is the only change, but it changes everything. And of course, once you think about it, it makes sense: if all that is in the title is a noun, then the noun must guarantee an interesting story all by itself, and vampires and parricides are a very good choice. (195)

An accompanying bar chart contrasts the presence of adjectives in two “semantic fields” (196, fig. 11). The semantic fields are not precisely specified, nor is a hypothesis formulated and tested. The argument rather derives its force from the inference to an explanation: “if you think about it, it makes sense.” The sense-making here is a compelling moment of expert reading. Its power stems from the move from grammar (nouns, adjectives) to meaning—the kind of move that typifies classic close reading. Moretti is explicit about his aim: “any new approach—quantitative, digital, evolutionary, whatever—must prove . . . that it can do formal analysis, better than we already do” (204). This “formal analysis” entails, in its rhetoric and logic, what is usually called close reading. Moretti addresses his linguistic facts as a text, assuming that those facts point to an integrated, significant whole.

The point applies even more strongly to the treatment of one of the signature features of computer-powered “distant reading” work: the use of data

visualization. Looking back on the work of the Literary Lab in a recent pamphlet, Moretti asserts:

Images come first, in our pamphlets, because—by visualizing empirical findings—they constitute *the specific object of study of computational criticism*; they are our “text”; the counterpart to what a well-defined excerpt is to close reading. (“Literature, Measured” 3)

As an account of a quantitative methodology, this is a strange statement: visualizations are powerful exploratory tools, but they should be considered provisional summaries of the data of literary history, not primary objects. As a description of argumentative rhetoric, however, Moretti’s analogy between visualization and excerpt is illuminating. It positions the “computational critic” as an expert interpreter of visual texts, a heterodox version of the close reader. “Network Theory, Plot Analysis,” the last essay in *Distant Reading*, closes with a plea for “sharing raw material, evidence—*facts*” (240). But that essay, like “Style, Inc.” and much of the work since, instead makes its case by interpreting data visualizations (bar charts, line charts, network graphs) as clues to a cultural totality. These are not facts but readings.

#### BEYOND READING (AGAIN)

The tension between the earlier and later parts of *Distant Reading* recurs in Moretti’s more recent work when he and his collaborators return to the question of sociology in a pamphlet, “Canon/Archive” (Algee-Hewitt et al.). This essay dramatizes the split possibilities of distant reading in its two halves: the first, an explicitly Bourdieu-inspired attempt to map “the British novelistic field” (4) by placing authors on two axes of prestige and popularity; the second, a “morphological” inquiry (12) which seeks to show that the “archive” of largely forgotten nineteenth-century British novels is markedly distinct from the “canon” on measures of lexical redundancy. The first inquiry, operating at the level of data about reception, is an aggregate of high-level responses by human agents to literary texts in the world, and it produces an image of the literary system which the pamphlet authors juxtapose to one of Bourdieu’s field diagrams from *The Rules of Art*. This analysis recalls distant reading in the mode of “Conjectures.”

The corpus study is distant reading in its later version, exemplifying the textualizing effects of the national framework and the corpus as object. The au-



thors insist that “the sociology of the literary field *cannot rest on sociology alone*: it needs a strong morphological component” (12). The exposition includes numerous interpretive readings of diagrams and of individual cases taken to be exemplary of canon and archive (9–10). The key claim is, as promised, “morphological”: “redundancy [is] apparently decisive in shaping the destiny of books; but the whole process took place so far below the level of conscious reading as to be practically invisible” (7). The rhetorical pattern is clear: positing a mechanism down where “conscious reading” does not reach, the pamphlet *assumes* the determining power of literary form (really, diction). This assumption, normal in expert literary reading, is in tension with the sociological study of the literary system. And its effort to supersede “sociology alone” is a matter of disciplinary prior conviction—which is to say, of doxa.

My own preference for the earlier version of distant reading is no doubt evident. Still, I have exaggerated the contrast for polemical purposes; the development of distant reading I have sketched was possible because the two varieties of distant reading joined in “Canon/Archive” lie on a continuum of possibilities. There were and are good reasons to turn to computation, to text corpora, and even to data visualization to answer significant literary-historical questions. The digitized archive is an extraordinary resource, and computer-aided analysis opens new lines of inquiry into literary systems and aggregates. Furthermore, as Underwood points out, the computational turn brings some “distant readers” into fruitful conversation with developments across the human and social sciences (“Distant Reading”). Nonetheless, the newer version of distant reading, despite its computational apparatus, is more a kind of literary-critical “reading” than the earlier distant reading was. If we cease to regard the different versions of distant reading as a singular project—even though Moretti, in his remarkable range, exemplifies them both, and more besides—we gain a wider sense of the possibilities for scholarship. Conflating the two, by contrast, makes it harder to study, rather than affirm, the organizing principles of the literary field.

#### WORKS CITED

- Algee-Hewitt, Mark, et al. “Canon/Archive. Large Scale Dynamics in the Literary Field.” Stanford Literary Lab, 2016.
- Best, Stephen, and Sharon Marcus. “Surface Reading: An Introduction.” *Representations*, vol. 108, no. 1, November 2009, pp. 1–21.

- Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *MLQ*, vol. 78, no. 1, March 2017, pp. 77–106.
- Bourdieu, Pierre. *The Rules of Art: Genesis and Structure of the Literary Field*. Translated by Susan Emanuel, Polity, 1996.
- Cecire, Natalia. "Ways of Not Reading Gertrude Stein." *ELH*, vol. 82, no. 1, 2015, pp. 281–312.
- Cohen, Daniel J., et al. "Interchange: The Promise of Digital History." *Journal of American History*, vol. 95, no. 2, September 2008, pp. 452–491.
- Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton UP, 1999.
- DeWitt, Anne. "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press." *Victorian Periodicals Review*, vol. 48, no. 2, Summer 2015, pp. 161–82.
- During, Simon. "Katherine Mansfield's World." *Journal of New Zealand Literature*, no. 33, 2015, pp. 33–66.
- Gold, Matthew K., and Lauren Klein, editors. *Debates in the Digital Humanities 2016*. U of Minnesota P, 2016.
- . Introduction to "Forum: Text Analysis at Scale." Gold and Klein, *Debates*, [dhdebates.gc.cuny.edu/debates/text/93](http://dhdebates.gc.cuny.edu/debates/text/93).
- Griswold, Wendy. "American Character and the American Novel: An Expansion of Reflection Theory in the Sociology of Literature." *American Journal of Sociology*, vol. 86, no. 4, January 1981, pp. 740–65.
- Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry*, vol. 42, no. 2, Winter 2016, pp. 235–67.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. U of Illinois P, 2013.
- Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. 3rd ed., Sage, 2013.
- McGill, Meredith L. "Echocriticism: Repetition and the Order of Texts." *American Literature*, vol. 88, no. 1, January 2016, pp. 1–29.
- McMillan Cottom, Tressie. "More Scale, More Questions: Observations from Sociology." Gold and Klein, *Debates*, [dhdebates.gc.cuny.edu/debates/text/98](http://dhdebates.gc.cuny.edu/debates/text/98).
- Mitrić, Ana. "Jane Austen and Civility: A Distant Reading." *Persuasions*, no. 29, 2007, pp. 194–208.
- Moretti, Franco. *Distant Reading*. Verso, 2013.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.

- . “Literature, Measured.” Stanford Literary Lab, 2016.
- Mufti, Aamir R. “Global Comparativism.” *Critical Inquiry*, vol. 31, no. 2, Winter 2005, pp. 472–89.
- Radway, Janice. *Reading the Romance: Women, Patriarchy, and Popular Literature*. U of NC P, 1984.
- Schulz, Kathryn. “What Is Distant Reading?” *New York Times Book Review*, 24 June 2011, p. 14, [www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html](http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html).
- Trumpener, Katie. “Paratext and Genre System: A Response to Franco Moretti.” *Critical Inquiry*, vol. 36, no. 1, Autumn 2009, pp. 159–71.
- Underwood, Ted. “Distant Reading and Recent Intellectual History.” Gold and Klein, *Debates*. [dhdebates.gc.cuny.edu/debates/text/95](http://dhdebates.gc.cuny.edu/debates/text/95).
- and Jordan Sellers. “The *Longue Durée* of Literary Prestige.” *MLQ* 77, no. 3, January 2016, pp. 321–44.
- Wilkens, Matthew. “The Geographic Imagination of Civil War-Era American Fiction.” *American Literary History*, vol. 25, no. 4, Winter 2013, pp. 803–40.