

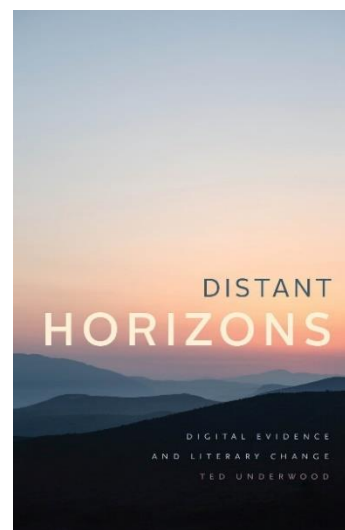
Ted Underwood

# Distant Horizons

Digital Evidence and Literary Change

Chicago, The University of Chicago Press, 2019,  
pp. 206

Review by Alice Crescini



**Keywords:** *Digital Humanities, distant reading, literary theory, review*

As could be easily grasped from its title, Ted Underwood's latest work *Distant Horizons: Digital Evidence and Literary Change* aims at unfolding a wider new perspective on literary history through digital means of research. Starting from broad questions and "historical arguments" (xii), it focuses on the analysis of specific case studies and leaves academic discussion of debatable issues in the background, albeit not overlooked. Rather than displaying and defending the legitimacy of distant reading's methods, Underwood tries to underline their parallel role to traditional procedures of inquiry in the broader field of literary theory:

[...] Distant reading is simply a new scale of description. It doesn't conflict with close reading any more than an anatomical diagram of your hand would conflict with the chemical reactions going on inside your cells. Instead of displacing previous scales of literary description, distant reading has the potential to expand the discipline—rather as biochemistry expanded chemistry toward a larger scale of analysis. (xvii-xviii)

This approach intends to shift from longstanding concerns of the discipline—such as those regarding Moretti's "slaughterhouse of literature" (2020, 51-68)—towards new perspectives which imply the involvement of long timelines of change and wide patterns in the study of literary history. The author dismisses the struggle on full archival recovery and states the necessity, to borrow Benjamin Schmidt's words, of understanding "a source *through* its biases" (2012): instead of seeking the unbiased sample of volumes, Underwood intends to identify and analyze the biases so as "to measure the parallax between different observers" (xv). This approach is called "perspectival modeling": "by training models on evidence selected by different

people, we can crystallize different social perspectives and compare them rigorously to each other” (xv). In this way, the machine learning techniques’ tendency to absorb latent assumptions from the sample they are trained on turns into a useful tool of comparison and analysis.

As the author states in the Preface, this work—despite relying upon a different and often controversial approach—is merely “[...] a book about the history of English literature, focusing especially on Anglo-American writers” (xii): the graphs do not crowd the pages and most of the discussion over data and methods is entrusted to two appendixes placed at the end of the volume. Thus, this book is easily readable by scholars not specifically trained in mathematics, social science, or statistics. It presents its arguments in a way that is clear, straightforward and effective at the same time, effortlessly guiding the reader through figures and models. Further data and all of the code employed are provided in an online repository (whose link is made available by the author), so that it can be tested and checked. The main source for digital texts used in this work is the HathiTrust Digital Library.

Starting from the fundamental assumption that “there are at least a few broad trends in literary history that we don’t yet understand” (14), the first chapter considers changes in 18th-, 19th- and 20th-century fiction and relates them to a long process of differentiation of the genre from that of nonfiction. Quantitative evidence to this is provided through a statistical model (i.e., a relation between predictor and predictive variables) which predicts the probability that a book will be perceived as fiction. The model used is logistic regression plus “regularization—a degree of deliberate blurriness that prevents the model from memorizing the examples in its training set, forcing it instead to produce a looser, more portable generalization” (21). The portability of the model is verified via cross-validation, which also tests its accuracy. The peculiarity of this model is that it focuses on both textual and social evidence, overcoming mere text measurement and going towards a more relational approach. The author avoids sharp correlation between aspects emerged in the research but stresses the importance of highly detailed description of trends—herein, the quantitative evidence of the striking separation of the narrative forms and themes of fiction from those of nonfiction across three centuries.

The second chapter delves into perspectival modeling and applies it to the history of genres. Genre is initially defined as “a group of books recognized by some specific, historically situated group of readers” (37). Starting from recent scholarly opinion, library catalogers’ practices over time and 19th-century reviewers, the author identifies and tests lists of titles in three categories—detective fiction, science fiction (in which he includes scientific romances), and the Gothic—in order to understand how stable those genres have been through the years, decades or centuries. Expecting to find “a story of gradual consolidation” (40), Underwood discovers

instead that science fiction and detective fiction display stable textual patterns over the same span of time, from the beginning of their circulation. Consequently, the author challenges the mainstream definition of genre stating that

a genre is not a single object we can observe and describe. It may instead be a mutable set of relations between works that are linked in different ways and resemble each other to different degrees. [...] our goal is no longer to intuit a definition but to find a model that can reproduce the judgments made by particular historical observers. (41)

Models can individuate a large quantity of values and relations that could be used to describe a genre; however, the author also admits that in order to define its essential features it is fundamental to open up a critical debate.

The third chapter focuses on the intent to trace the emergence and changes of literary prestige from the 19th century to the 20th, defined as “a widening gulf between mass culture and elite literary taste” (70) in Andreas Huyssen’s “great divide” (1986). This chapter’s findings confirm preexisting affirmations that growing audiences in the Victorian era led to a gradual diversification of the market. The methods used for this part of the research are analogous to those used in the previous chapter (word frequency count, predictive modeling, machine learning) and have been applied to reviewed and random works of poetry and fiction in order to measure reception. The research highlights that literary production and the criteria of its judgment generally tend towards prestigious examples, but wisely chooses to avoid attempts to define precise indicators of literary prestige. Shifts in the results due to “algorithmic bias” in gender and nationality are also inspected; moreover, the model was tested adding information regarding sales and therefore meaning to measure popularity and contrast it with the data concerning prestige. Overall, the results underline a general pattern which is not particularly coherent or clear but instead consists of “several different axes of distinction” (95).

The fourth chapter explores the history of characterization, highlighting implicit suppositions concerning gender roles in English-language fiction from 1780 to the present. This chapter makes use of the software tool BookNLP, developed by David Bamman, which has some blind spots when drawing the boundary of identity (i.e., characters referred to with unusual nicknames might be split into multiple roles) but is quite precise in assigning gender roles. In order to trace the implicit assumptions that define them, characters are represented by the adjectives that modify them, the action verbs they use and other markers, excluding words that explicitly designate a gendered role. The algorithm learning on those characters can produce predictive models. Underwood describes this part of his research as harder to set out than those

illustrated in previous chapters: the methods used here (i.e., natural language processing) are more complex and the evidence is often blurry, its accuracy ranging from 64% to 77%.

Having discussed and argued the main points of the research, the author gives some space to the exploration of “The Risks of Distant Reading”: as per its title, the fifth and last chapter is intended to investigate major controversies over the adoption of these methods of inquiry in literary history. Restating its main points, the author states that “the thesis of this book is that hasty assumptions about narrative interest have focused literary history too narrowly on the scale of the individual author or generation, leading us to ignore spectacular longer-term trends” (148). The author claims that distant reading is not meant to—and in any case should not—overcome traditional approaches of analysis; however, its findings indeed have the potential to broaden the discipline. Underwood also discusses the impact of introducing numbers and statistics in the humanities and the way in which this can work:

[...] we can never afford to let the sophistication of our methods upstage the human interest of our subject. [...] Distant readers will need to practice an unobtrusive kind of quantitative rigor—scrupulous and patient with details but partly deferred to appendixes so the literary subject itself can take center stage. (149)

Finally, the author discusses what in his opinion is the main barrier to a full adoption of distant reading in the academic context: the curricular foundation of most literary historians does not include social sciences, statistics or computer programming. This problem is harder to solve, being essentially institutional.

To conclude, with this work Ted Underwood challenges mainstream decadal and generational ways of depicting literary change in favor of long timelines and wider perspectives. However, as the author states, with distant reading “we may have acquired enough altitude now to see that the horizon of literature is curved. But that discovery should only remind us how little we understand literary history as a whole” (169). This book asks many questions: some of them are answered, some will need further research. Underlining the potential of distant reading in uncovering new objects of knowledge and perspectives, the author also states that the field is not intended to displace close reading and the patterns uncovered are not alternatives to the canon. Most of the relevant discoveries, such as the continuous—rather than conflictual—pattern of change in literary movements, are descriptive. Interpretation and critics are still required for a full understanding of literary history, although it is quite clear that the methods mentioned here could be of much help—if not eye-opening.

**Alice Crescini** is a master's degree student in Publishing and Journalism at the University of Verona. She obtained her bachelor's degree in Languages and Cultures for Publishing in 2020. Her final paper featured among the activities of the PRIN 2017 project (Mapping Chivalry. Spanish Romances of Chivalry from Renaissance to 21st century: a Digital Approach). She began her collaboration with Iperstoria as an intern and is now a member of its editorial staff.

### Works cited

Huyssen, Andreas. *After the Great Divide*. Bloomington: Indiana University Press, 1986.

Moretti, Franco. "Il mattatoio della letteratura." *A una certa distanza. Leggere i testi letterari nel nuovo millennio*. Edited by Giuseppe Episcopo. Rome: Carocci, 2020. 51-68.

Schmidt, Benjamin. "Reading Digital Sources: A Case Study from Ship's Logs." *Sapping Attention*. November 15, 2012. <http://sappingattention.blogspot.com/2012/11/reading-digital-sources-case-study-in.html>. Last visited 10/06/2021.