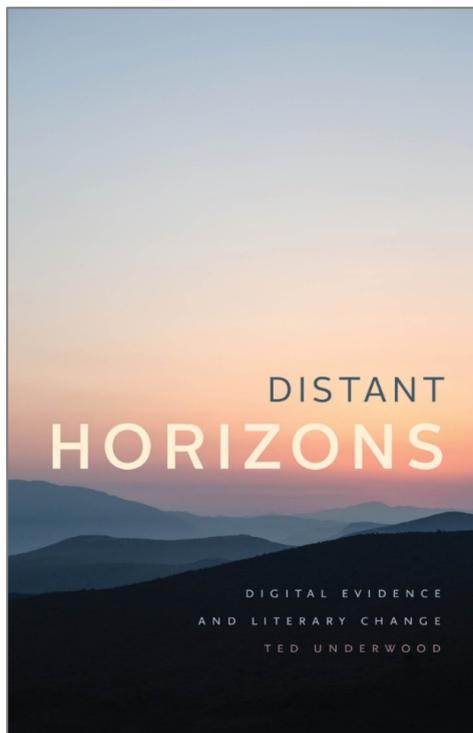


BOOKS

Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change*, Chicago: University of Chicago Press, 2019, 200 p.

Could distant reading ever work? Could digital archives and statistical tools deepen our understanding of issues that have always been central to humanistic inquiry? These are the very premises of Ted Underwood's latest book, *Distant Horizons*, published in 2021 with University of Chicago Press. This book contains a completely fresh and original approach to literary theory, accessible to both experts and students in the humanities. It shows how digital

methods can bring into focus the bigger horizon of literary theory and add to the beauty and complexity we value in literature, as the afterword highlights. The book is structured in five chapters, each of them addressing a highly topical issue: 1. *Do We Understand the Outlines of Literary History?*, 2. *The Life Spans of Genres*, 3. *The Long Arc of Prestige*, 4. *Metamorphoses of Gender*, 5. *The Risks of*



Distant Reading. But as Underwood mentions in the foreword, the methods he will be describing do, of course, have limits. The author's mention makes the work an assumed and conscious project, open to criticism and questioning. What is even more fascinating is the fact that he writes in a way that is not difficult to process—he reaches a “delicate balance,” writing about interdisciplinarity without going into too much detail. In other words, the technical details of his methods and approach are explained in an easy-to-understand, but by no means simplified way. Moreover, throughout the book, he uses logistic regression to model the data, producing elegant graphs and (hopefully) reproducible results.

Describing dimensions of literary history, such as reception and genre, in which volumes can be discussed as wholes,

the first three chapters state that topics like plot and character are harder to trace across long timelines because they require divisions below the volume level that are challenging to tease out algorithmically. Thus, with collaborative support from computer scientists, it is also possible to make some progress on those topics (17). The work described here owes something to twentieth-century projects like book history, stylistics, and the sociology of literature, as well as to the more recent fusion of those projects that goes under Franco Moretti's term "distant reading" (11).

The first chapter is, I think, one of the strongest pleas for distant reading, suggesting that that many well-known changes in eighteenth-, nineteenth-, and twentieth-century fiction can be understood as parts of a single differentiating process that defined the subject, style, and pace of fiction through opposition to nonfiction (14). As Michael Falk writes in his own review, Underwood's incredible humility is attractive, in a field where *Wunderkinder* often make extravagant claims about their digital research, and invent silly mystical-scientific names for their normally rather mundane methods—Underwood really cares about getting the right answer. Moreover, he studies how a wide range of artistic movements, often said to conflict with each other, sometimes said to have sought rapprochement with "ordinary language," have all actually pushed fiction farther away from the language, themes, and narrative strategies of nonfiction (14).

Chapter 2 focuses on how these new methods can support a perspectival approach to genre: "Genres are not the only human creations that change their meanings with time. The interpretive problems

that confront a history of genre are rooted in the perspectival dimension of history itself, and they run too deep to be solved neatly" (15). Furthermore, the term *science-fiction* is brought up, suggesting the fact that its meaning will depend on an observer's location. This may signify that *science-fiction* itself can have more interpretive resonances as time goes on. And that is exactly where the mathematics aspects come in, because it is all about questions of perspective. The so-called "machine learning," says Underwood, is constantly causing public scandal due to its tendency to be all too sensitive to subjective contexts. This is exactly the reason why, continues the author, institutions that strive to be unbiased might well choose to avoid machine learning (16). Actually, Underwood theorizes a new approach which he calls "perspectival modeling":

Readers who are familiar with other ways of using machine learning may need to set some assumptions aside. The models created in this book are supervised: that is, they always start from evidence labeled by human readers. But unlike supervised models that try to divine the real author of an anonymous text, perspectival models do not aim simply to reproduce human judgment. They are used instead to measure the parallax between different observers (16).

The third chapter, *The Long Arc of Prestige*, explains how questions of form and genre intersect with grittier aspects of literary production and distribution. For Underwood, textual forms are important aspects of literary pleasure: there is no reason to apologize for studying them. But they are not the only part of history that can be enriched by quantitative

reasoning (68). Until recently, in fact, numbers were far more useful for social questions than for aesthetic ones. That means that unemployment, inequality, and book sales are easy to measure. But it is not intuitively obvious how one would measure literary style. In making it possible to construct quantitative models of genres and styles, machine learning has made it easier to link formal concepts to quantitative social evidence and thus to build bridges between social history and the history of texts (69).

The thesis of this chapter is that a different kind of description is possible, which will combine the rigor of a detailed account with the ambitious scope of a larger narrative. When we train a model using social and textual evidence across a whole century, we can describe long-term patterns that connect social pressures to persistent directions of literary change. This, however, does not make shorter-term trends unimportant, but it puts them in a different perspective. As an example, Underwood portrays the following situation:

Imagine if we could show, for instance, that food had been getting steadily spicier in the United States for the past century and that the best-reviewed restaurants had consistently occupied the leading edge of this trend. The stories we tell about decade-long culinary trends might still be true. But those stories would also have to be seen as parts of a broader pattern, which would become central to any explanation of long-term culinary change (70).

Also, the third chapter argues that something analogous is true about English-language poetry and fiction between 1820 and 1949. Our received narrative of

this period is organized by a succession of discrete concepts defining different criteria of judgment: Romanticism, Victorian realism, aestheticism, naturalism, and modernism. For many observers, this has implied a fairly profound transformation of literary opinion every generation or so. Modernism, for instance, was a “literary revolution” that changed not only how writers created new works but how they evaluated the past, producing “a radical and wholesale revision of the inherited conception of English literature,” as Chris Baldick shows in *Modernist Criticism and the English Literary Canon*.

Moving forward on this journey, chapter four explores the history of characterization, looking in particular at the way fictional characters are shaped by implicit assumptions about gender. Underwood goes into some detail about the strategies he used to validate his models, and analyzed a whole series of examples to try and explain how his model related to the reality that it modelled. Statistical tests and data tables make clear exactly what had been modelled and how. He insists that historians of literature obviously need a bit more distance from fiction (102), connecting at the same time our history to readers’ experience, which means that we need some way of connecting historical trends to the imaginary people and events inside the volumes. To put it in simple words, we need some way of reasoning collectively about hundreds of thousands of fictional people.

This chapter takes a few steps in that direction, in order to trace the history of gender roles in English-language fiction from 1780 to the present. But even a few steps toward a history of character will admittedly take us to the

edge of what is now possible. The methods discussed here are more complex than those discussed in earlier chapters, the evidence often a little noisier (112).

Chapter five concludes the book with a defense of 'distant reading' that perhaps explains why Underwood adopted this unique style for the book. The author expresses some anxiety that putting too many numbers into a work of literary history will turn literary colleagues away, and says at one point that a technical appendix is probably the best place to add the statistical information. In the first decade of this century, many people hoped that quantitative methods could be introduced to the humanities in an equally painless way. Underwood thinks that digital humanists would build tools and that everyone else would use them. Scholars might not need to understand all the details inside the box, any more than we have traditionally worried about the innards of a search engine when fishing for sources:

Foundations invested millions of dollars trying to support this quick, painless kind of change. But for the most part, it didn't happen. Instead, change has taken place slowly, and mostly through laborious retraining. The reason, I think, is that new methods have turned out to be more consequential than was widely believed a decade ago. Search engines can be encapsulated and treated as tools. But statistical models are not well envisioned as tools: they offer new methods of representing and interpreting the world. Scholars cannot adopt a new mode of interpretation without fully understanding the reasoning it implies (145).

To put it in a nutshell, Ted Underwood's *Distant Horizons* is an excellent introduction to the possibilities of quantitative literary history, computational literary studies and distant reading. Underwood's prose is clear, the case studies are interesting and the use of computers to analyze the character traits of novels is a fresh and fascinating method.

MARIA BUCŞEA

Independent researcher

Email: maria_bucsea@yahoo.com