The Digital Humanities and Literary Studies

Brief Synopsis

This short book delivers an introduction and overview of developing intersections between digital methods and literary studies. The volume will serve as the best starting place for those who wish to learn more about the possibilities, but also the limitations, of the oft-touted digital humanities in the literary space. The volume will engage with the proponents of digital humanities and its detractors alike, aiming to offer a fair and balanced perspective on this controversial topic. The book combines a survey and background approach with my own original literary research and is, therefore, well-placed to straddle the divide between seasoned digital experts and interested newcomers.

Biography

Martin Paul Eve is Professor of Literature, Technology and Publishing at Birkbeck, University of London. Martin holds a Ph.D. from the University of Sussex and is the author of five other books: *Pynchon and Philosophy: Wittgenstein, Foucault and Adorno* (Palgrave, 2014); *Open Access and the Humanities: Contexts, Controversies and the Future* (Cambridge University Press, 2014); *Password* (Bloomsbury, 2016); *Literature Against Criticism: University English and Contemporary Fiction in Conflict* (Open Book Publishers, 2016); *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas* (Stanford University Press, 2019). He has held three large grants from the Andrew W. Mellon Foundation and before beginning his academic career, Martin worked as a computer programmer.

Past Endorsements

"[*Close Reading with Computers*] marks a major intervention in debates about digital humanities, literary criticism, and textual scholarship. It demonstrates commanding knowledge of work in these broad fields, and provides persuasive perspectives on several interrelated questions taken up singly by other scholars. […] I do not know of any other scholar who has such facility with computer reading and computational methods AND such a subtle literary critical sensibility. This is truly a unique work." – Professor Paul Harris, Professor of English, Loloya Marymount University

"Martin Paul Eve is one of the most brilliant scholars of his generation. His ground-breaking *Literature Against Criticism* combines new and insightful readings of contemporary novelists (from Jennifer Egan to Tom McCarthy and from Sarah Waters to Percival Everett) who are in animated competition with university English. There are very few authors who can combine ethical, political

and aesthetic readings of the contemporary novel with an encyclopaedic knowledge of the modern university.” – Bryan Cheyette, Professor of English, University of Reading

“Conjuring our passwords has become a daily act of our computer-saturated existence. By no means sequestered to our digital present, Martin Paul Eve's excellent account of the password covers its long and lively history. Weaving literary references with lucid technical explanations, Eve skillfully traces the evolution of password to probe its fundamental connections to issues of human identity, trust, and ownership.” – Gabriella Coleman, Wolfe Chair in Scientific and Technological Literacy, McGill University, Canada

"Martin Paul Eve's *Pynchon and Philosophy* is a work of consummate scholarship. Breaking new ground in Pynchon studies, Eve offers an immensely erudite, detailed and in-depth account of the ways in which the ideas of Wittgenstein, Foucault and Adorno help us to think about his texts. A first-rate book." - David Cowart, Louise Fry Scudder Professor of Humanities, University of South Carolina, USA

Proposed Table of Contents

Projected extent: 55,000 words.

Chapter-by-Chapter Summary

I: Introduction: Words that Count [10,000 words]

The study of literature with the aid of computers is controversial. Digital methods in literary studies have been derided for being useless (“they tell us nothing that we did not already know”), neoliberal (“producing software is the Silicon Valley model of scholarship”), and trivial (“counting the word whale in *Moby Dick* can tell us only one thing: how often the word whale is used in *Moby Dick*”). Proponents on the other hand have pronounced forcefully on the possibilities for broad-scale literary history beyond the limitations on reading made by the limited human life span; on the ways in which we can better understand genre and form through visualization and spatialization; and even on the fresh perspectives such methods might bring for rethinking core theoretical assumptions about literature itself.

This introductory chapter serves to frame the volume, sequentially addressing the various claims made for digital approaches to literary studies. It begins by noting that there is, in fact, a rather long history of quantitative methods (which underpin all digital methods) in the space of literary analysis. For instance, Thomas Schaub claimed in 1981 that "the word 'bloom' is one of the most oft-repeated words in" Thomas Pynchon's *The Crying of Lot 49* (although he is actually incorrect), while Dartmouth College offered a course entitled "Literary Analysis by Computer" as far back as 1969. Further, as Nicholas Dames has noted, Vernon Lee proposed a "statistical experiment" – a quantitative analysis – on literature in her 1923 *The Handling of Words*, itself prompted by a letter to *The Times* from Emil Reich, several years earlier.

Yet, the application of computational power to such quantifying problems foregrounds a set of anxieties in the space of literary studies, to which the remainder of this chapter devotes itself. In particular, I turn to address the discursive constructs of digital humanities, as Matthew Kirschenbaum has framed it, rather than particular projects. These include the critiques of the utility of digital humanities, its situation as "neoliberal" (Golumbia, Allington, and Brouillette), and longer philosophical objections to treating the work of art as the subject of quantitative or scientific investigation (such as the first generation Frankfurt School).

By the close of this chapter, the reader should understand the discursive field within which the "digital humanities" finds itself positioned, while also having developed a grasp of the promises held out by new digital approaches, to which the subsequent chapters then turn.

II: Authors and Writing [10,000 words]

Two central questions posed by literary theory over the past half century have been: "what is a literary text?" and "what is an author?". Indeed, the university discipline of literary studies, ever since its formal inception in 1828 at University College London, has never truly known its precise object of study. What might it mean that for a text to be particularly "literary"? Is there any discernable aspect within language itself that denotes a work as literary? Certainly Jacques Derrida and John Searle, among others, believed not. Yet, as Andrew Piper has recently shown, neural networks are able to discern fiction from non-fiction, with about 95% accuracy. To be clear: these networks are not verifying the truth claims of non-fiction against fiction; they are assessing the formal characteristics of these two forms of writing.

In this chapter, I introduce a range of approaches to the measurement and digital quantification of literary style: stylometry. This begins with a history of stylometric thinking, ranging from approximately 1851 when Augustus de Morgan suggested that a dispute over the attribution of certain epistles could be settled by measuring average word lengths and correlating them with known writings of St Paul, through to contemporary multi-dimensional fingerprinting techniques, such as Burrows's delta method. I then progress to discuss questions of close vs. large-scale literary reading and the problematic terminology of "distant reading" (namely, that one can use computational techniques also to read closely, despite this also being a type of "distant" reading).

Also covered in this chapter are analyses of genre and its lifecycles (Ted Underwood), approaches to poetry (Tanya E. Clement on Gertrude Stein), court cases that have turned on stylometry, the

Mosteller and Wallace methods applied to the Federalist papers, and authorship controversies (Don Foster's infamous digital work on Shakespeare).

I close this chapter by returning to my own conceptual work on the assumptions behind authorship and literary language on which stylometry hinges, first outlined in *Close Reading with Computers*. I here argue that most stylometric approaches share a number of common underpinnings: 1.) that authors have a "stylistic naturalism"; 2.) that stylometry measures subconsciously inscribed features of a text; 3.) that authorship is the underlying textual feature that can be ascertained by the study of quantified formal aesthetics. I end by rehearsing the implications of such thinking upon the two literary-theoretical questions with which I began this chapter.

III: Space and Visualization [10,000 words]

The common link between the section titles of Franco Moretti's book, *Graphs, Maps, and Trees* (2007), is the visuality of his abstract models for literary history. For graphs, maps, and trees are all structures by which we can downmix complex, multi-dimensional aspects of literature into approximate two-dimensional (or sometimes three-dimensional) space. Much like conventional literary criticism, visualization yields to us new ways to conceive of narrative, re-orienting texts through fresh optics and augmenting understanding. Visualization is a form of deformance and interpretation, as Jerome McGann and Lisa Samuels would have it.

In this chapter I examine a range of approaches to visualization of literary works, picking up where Chapter II left off with visualizations of authorship attribution. In particular, I relate this specific technique to the underlying statistical certainty of the claims being made and present an explanation of bootstrap consensus trees as a potential solution (as suggested by Eder). From here, I quickly move to the textual clustering maps developed by David McClure to examine Tolstoy's *War and Peace* (1869) and Thomas Pynchon's *Gravity's Rainbow* (1973), demonstrating the ways in which computational visualization approaches can assist in navigating dense, interrelated works. Finally, for the methods in this chapter, I turn to the ways in which it is possible to visualize textual genetics and textual criticism. This returns to my earlier work on *Cloud Atlas*, but also presents fresh research into the early work of the Pulitzer prize-winner, Jennifer Egan, showing for the first time that her earliest story collection was published in two very different editions, and using computational visualizations to illustrate the differences.

By the close of this chapter the reader should have a grasp of how and why visualization might be a helpful technique, while also understanding its limitations as a reductive practice that nonetheless aids understanding.

IV: Place and Maps [10,000 words]

The "spatial turn" in the humanities – exemplified in the work of scholars as far apart as Jo Guldi and Robert Tally Jr. – draws our attention to the way in which literary texts structure their senses of place. From J.R.R. Tolkien to W.G. Sebald via the Hundred Acre Wood, literary works have often, also, included maps within their pages. Yet such *topoi* sit distinct and apart from the extra-textual world, even when such places are represented therein.

Digital approaches to geographic information systems (GIS) have been among the most commonly deployed technologies to think "around" these issues of space and place. Whether it be in visualizing the multiple pathways taken by Woolf's characters in *Mrs Dalloway* or mapping the Lake District of the Romantic poets, attention to literary geography has been extensive in the digital world.

This chapter introduces the reader to the variety of approaches that have been undertaken in the GIS space, beginning with a theoretical background to ideas of literary space before moving to a showcase of various projects and their accomplishments. Drawing on the broad intellectual history of space and literature, from Freud's influence on mental space, through to Bachelard's influential *The Poetics of Space* (1958), via Walter Benjamin's *Arcades Project* (2002), I here appraise the ways in which digital models of geography have had to accommodate the difficult interrelation between literature, representation, and reality.

Finally, this chapter also features new research work on mapping approaches to Mark Blacklock's 2015 novel, *I'm Jack*, which centres on the Wearside Jack hoaxer and the telecommunications paradigms that surround geolocation in that text.

V: Distance and History [10,000 words]

For many years now, more contemporary fiction has been published every year than it is possible for a single person to read in a lifetime. In 2015, according to Bowker data, almost three million new books were printed in English alone, of which two hundred and twenty thousand were novels. A good estimate for the number of days in a human lifespan is 26,000 (approximately 71 years), using the World Health Organization's figures as of 2015, so one would need to read an average of ten novels per day, every day, from age ten onwards, to have read all English fiction published in 2015.

The implications for literary history here are enormous. Field mastery by a single individual is impossible and the systematizing dreams of the early Russian formalists seem far out of reach. One of the ways in which statistical reading has been billed as useful, though, is in overcoming these human limitations. If we cannot read enough ourselves, perhaps, it is posited, we might delegate this work to the machines.

In this penultimate chapter, I explore the advances in our understanding of large-scale literary history advanced by figures such as Ted Underwood, Andrew Piper, James F. English, Andrew Goldstone, Jessica Pressman, and Matt Jockers. The chapter tackles the relationship between the small-scale literary modelling required to refine computational methods and its scaling across vast corpora such as the HathiTrust archive. This chapter also broaches the legal issues of copyright and non-consumptive use paradigms.

VI: Conclusion: Politics and the Personal [5,000 words]

Although it is impossible to provide a totalising summary of an area as broad as "the digital humanities and literary studies" in a book of this length, the final chapter here is devoted to an

understanding of some of the overlooked areas. In particular, I give the final words in this volume to feminist and postcolonial perspectives on digital humanities, and particularly the work of scholars such as Roopika Risam and Alex Gil. Far, then, from showing DH to be a complicit, neoliberal space, I end with the explicit ways in which the political work of our time is being conducted in institutions, using digital tools, thereby complicating arguments about the political alignment of digital practices.

Delivery Date

The manuscript can be ready by the 1st February 2020 or one year after greenlighting of the project, whichever is later.

Figures and Tables

The manuscript will have approximately 15 figures that can be customized for black-and-white printing. There are also several tables (approximately 6 at present).

Fit with Existing Literature

This work performs a different function to other digital humanities (DH) monographs on distant reading. It serves, twofold, as a specific exploration and showcase of digital methods in literary studies (rather than the broader field of the digital humanities more generally), but also as a space of new research; a hybrid work between an introduction and a research monograph.

The work is also aligned with, but differentiated from, the forms of computational poetics and thinking, such as that recently outlined by Dennis Tenen in his *Plain Text: The Poetics of Computation*. In one sense, the work does take much of the theoretical thinking from the social DH and New Media spaces (Alan Liu, Stephen Ramsey, N. Katherine Hayles, Susan Schreibman and others). In particular, I am indebted here to the work of Tanya E. Clement, who has done more than many to interrogate the relationship between close reading and digital practices.

This book comes at an apt time; there has been an explosion of interest in digital literary methods in recent days, with books by Ted Underwood (Chicago), Andrew Piper (Chicago), and myself (Stanford) having been published or are forthcoming within this season's catalogues. While there are many works that contest the definition of the "digital humanities" (as just one example, Fagerjord and Berry's *Digital Humanities: Knowledge and Critique in a Digital Age* sits in this space), a book such as this that will quickly orient the new reader while also demonstrating both possibilities and limitations will enable a broader field to decide for themselves on whether they wish to adopt new digital approaches in their study. I also anticipate that this volume would be suitable as an introductory text for the ever-expanding number of courses in digital humanities that are cropping up, worldwide.

Intended Audience

The intended audiences for this book are:

- Literary scholars interested in digital methods
- Contemporary literary scholars interested in textual scholarship and version variance
- Literary scholars interested in the spatial turn
- Digital humanists interested in the new research presented in this book
- Students of any of the above