

Preface

This book is part of a longer exploration of the relationship between technology and reading, one that has occupied me for most of my career. In *Dreaming in Books*, I studied how romantic literature made sense of the bibliographic reorganizations that were sweeping across Europe and North America at the turn of the nineteenth century. Romanticism was, in this reading, a movement deeply invested in understanding material and technological changes that we are in many ways still grappling with. *Book Was There* sought to understand the recent technological upheavals around books by paying attention to more embodied dimensions of reading. Whether it is the touch or sight of the page or the places and practices of note-taking, game-playing, sharing, storing, or consuming books, I wanted to show how these experiences differ profoundly between print and digital media. Finally, *Interacting with Print* turned to the ways historical actors engaged with their reading material to produce new kinds of social communities, new models of creativity, and new structures of knowledge. Written with twenty-two coauthors, *Interacting with Print* put theory into practice in an elaborate process of scholarly interactivity of its own.

This then is the intellectual background to the book you are reading. *Enumerations* is about how computation participates in the construction of meaning when we read. It argues that data and computation unquestionably have a role to play in understanding literature, but that the way we have so far approached this problem rests on a number of flawed premises. The notions of distance, bigness, or

objectivity that are largely in circulation right now rely on overly binary models of reading, largely untethered from past practices. *Enumerations* tries to show how these frameworks do not adequately capture the nature of computational modeling and its place within the rich history of reading. We still do not have a clear picture of how emerging quantitative methods speak to the questions that matter within the discipline of literary studies. This book is an attempt to align new kinds of models with old kinds of questions.

As I began to think about why I was interested in the question of literary quantity, I realized that it marked an even more general continuum with previous concerns. It belonged to my abiding interest in understanding the commensurability of seemingly incommensurable things. Instead of exploring the relationship between words and their objects, or bodies and reading, or paper and electronic books, as I had done previously, here I have moved to the relationship between letter, number, and image (in the form of the diagram). However disparate, behind each of these efforts lies the idea of translation, the act of moving between languages, cultures, and mentalities, as a core practice, but also an ideal, for humanistic scholarship. The first book I ever published was a translation, and it occurs to me now that I have been writing under this sign ever since. In the back of my mind, I keep trying to imagine an alternative future where students are not dutifully apportioned into silos of numeracy and literacy, but are placed in a setting where these worldviews mix more fluidly and interchangeably.

As much as this book represents a continuum, it also marks a breaking point, both from my past work, but also in the sense of something being broken. The research for this book began concretely when I started “retraining” myself in the field of computational text analysis several years ago, combining the practice of computer programming with that of quantitative reasoning. As hard as this process was and continues to be, computation allowed me to gain two fundamental insights about our discipline that had so far been overlooked. The first is the pervasive quality of textual repetition. The vast bulk of any single text consists of elements that repeat themselves with great frequency. These repetitions in turn multiply out in the world, giving coherence to entire domains of writing, such as genres, periods, modes, topoi, and careers. And yet, we have had no way of accounting for this fact of recurrence. It was as though we had elected to orient ourselves around rare events to protect ourselves from the vast majority of textual features (not to mention texts themselves). Focusing on a single dash in Heinrich von Kleist makes

sense only if you pretend that there are not tens of thousands of other ones floating around.

The second problem is what I discuss in the introduction as a science of generalization. Until recently, we have had no way of testing our insights across a broader collection of texts, to move from observations about individual novels to arguments about things like *the novel*. And yet, we make these generalizations all the time. Indeed, one could argue that generalization is a crucial aspect to any scholarly method. It is what allows us to identify the significance of a particular instance as well as the social and historical significance of some larger set of practices. It is how we move between part and whole.

As recent research has begun to suggest, those wholes have been expanding for some time. The scale of our categories (world literature, new media, post-canon) has been matched by an increasing attention to social critique, to questions of worldly “mattering.” And yet, our methods have remained largely unchanged. I will never forget the moment when I realized that the usual answer our field offers to initiates when faced with this problem—read more!—suddenly seemed incredibly, even senselessly, insufficient. As the Enlightenment scholar Johann Hamann once said, the imperative to read more feels like the punishment of carrying water through a sieve meted out to the daughters of Danaus. More reading could never by itself provide the evidentiary foundations to make categorical arguments—whether about Romanticism, modernity, the book, the novel, or even literature. We require some way of traversing scales, of testing our individual insights and observations against a set of texts that is more representative of the category about which we are speaking, especially as our analytical scales keep expanding. It is clear that *more* needs to be replaced by something we might call *method*. I had seen the crack in the table.

I take this expression from a short 16 mm film produced by the artist Paul Sietsema (“Anticultural Positions” [2010]). In it, we see a number of still photographs of the surfaces upon which he worked while making the photographs and paintings used in an earlier film (“Figure 3” [2008]), one that was itself largely concerned with the representation of surfaces, like paper and pottery. What makes these films so moving is Sietsema’s attention to the fissures and lines that corrugate any surface when looked at closely, the way he sees the furrows of surface. At one point, he shows us a marble tabletop in his studio, in which we see a slight crack. Behind the pristine surfaces of knowledge, the foundations upon which something else is made, Sietsema reminds us that we also

need to see the cracks, the places of vulnerability within the whole. It was these cracks in the otherwise smooth surface of reading that computation had allowed me to see.

Sietsema's imagery offers a useful metaphor in another sense, too, because it draws attention to the visibility of the materials we use when we create. Filming the surfaces upon which the objects of a film are made highlights the infrastructures of how we know things. As I argue in the introduction, one of the affordances of computational reading is the way it makes the critical project more legible than has traditionally been the case. While there will always be tacit dimensions to knowledge (as Michael Polanyi was the first to remind us), computation can be far more exo- than endoskeletal when compared with inherited critical practices. It is in this spirit that I have tried to make as much of the data and code used in this book publicly available. This includes over 7,000 lines of code (paltry for some, elephantine for me), as well as hundreds of tables of derived data from the primary data sets, all organized by chapter. While many of the primary data sets, which are described more fully in the appendix, cannot be shared, due to copyright, I provide code and tables of metadata about the collections for you to extract and build your own versions of them (or at least understand what has been included in them). I am trying to set a standard of reproducibility that will, I hope, gradually become more of a norm.

Throughout, I have adopted the convention of describing each model or calculation referenced in the text in the notes. I generally favor plain-language descriptions of models over formulas and equations. The notes also contain a subsequent reference to the accompanying piece of code beyond the book (i.e., see script 1.1), where the full implementation of a model can be reviewed in greater detail. In doing so, I am trying to strike a balance between the conventions of the humanities, which emphasize reading as a form of knowledge in its own right, and those of more quantitative disciplines, which put all the formulas and tables up front. Others may want a different approach, but my hope is that this allows for a thoughtful reading experience as well as the ability to replicate a model. It maintains the spirit of the foot- or endnote as a paratextual space with a difference. You are free to use the code for your own purposes or to try to reproduce the results I put forth here. I make no claims to elegance in programming, but I am confident that the scripts *work*, at least as of today. Durability has taken on a whole new scale of meaning when seen against the long timescales of bibliographic preservation.

At its heart, this book is an attempt to bridge two very different intellectual worlds and ways of thinking and reading. It would not have

been possible without much generosity on the part of people from both of these worlds, some of whom I explicitly name in the acknowledgments or notes, but there are many more. The field is too diverse to be captured by a single proper name. We would do well to acknowledge that. Throughout the research and writing of this book, I have received tremendous amounts of help from others. This work is unquestionably more collective than traditional scholarship in the humanities. But it is also more bootstrapped, to use a computing term borrowed from the world of horses, in the sense of being more improvisational. Much of what I have learned has been acquired through the meandering and chance encounters of someone making his way through new terrain. As Adam Hammond has argued, there is a DIY quality to programming and computational criticism that is inspiring and pedagogically encouraging. This book wants to convince you that if you are not already doing so, then you too can enter into the world of computational reading.

If we are going to foster this sense of exploration (and the potential for getting lost), then we ultimately need a more flexible model of what it means to be an expert. Alongside the expertise of specialization, we need to value the expertise of synthesis and mediation, what it means to speak two different languages, or codes, or embody two different mentalities simultaneously. This book is dedicated to all of those people who don't feel at home inside something, whether it is a culture, a club, or a discipline, and instead who think there is something important to be discovered, something novel and consequential, in the spaces between.

