

Digital Humanities as a diverse field of practices associated with computational techniques and reaching beyond print in its modes of inquiry, research, publication, and dissemination. This perspective is in line with that adopted by the NEH Office of Digital Humanities (Boble, Rhody, and Serventi, 2011), whose personnel were advised at the program's birth to define the Digital Humanities as broadly as possible. For my purposes, the Digital Humanities include, among other kinds of projects, text encoding and analysis, digital editions of print works, historical research that recreates classical architecture in virtual reality formats such as *Rome Reborn* and *The Theater of Pompey*, archival and geospatial sites, and, since there is a vibrant conversation between scholarly and creative work in this field, electronic literature and digital art that draws on or remediates humanities traditions.

N. Katherine Hayles, *How We Think* (2012), pp. 27-31

Scale Matters

Perhaps the single most important issue in effecting transformation is scale. Gregory Crane (2008a) estimates that the upward bound for the number of books anyone can read in a lifetime is twenty-five thousand (assuming one reads a book a day from age fifteen to eighty-five). By contrast, digitized texts that can be searched, analyzed, and correlated by machine algorithms number in the hundreds of thousands (now, with Google Books, a million and more), limited only by ever-increasing processor speed and memory storage. Consequently, machine queries allow questions that would simply be impossible by hand calculation. Timothy Lenoir and Eric Gianella (2011), for example, have devised algorithms to search patents on radio frequency identification (RFID) tags embedded in databases containing six million five hundred thousand patents. Even when hand searches are theoretically possible, the number and kinds of queries one can implement electronically is exponentially greater than would be practical by hand.

To see how scale can change long-established truisms, consider the way in which literary canons typically function within disciplinary practice—in a graduate program that asks students to compile reading lists for the preliminary examination, for example. Most if not all of these works are drawn from the same group of texts that populate anthologies, dominate scholarly conversations, and appear on course syllabi, presumably because these texts are considered to be especially significant, well written, or interesting in other ways. Almost by definition, they are not typical of run-of-the-mill literature. Someone who has read only these texts will likely have a distorted

sense of how “ordinary” texts differ from canonized works. By contrast, as Gregory Crane (2008b) observes, machine queries enable one to get a sense of the background conventions against which memorable literary works emerge. Remarkable works endure in part because they complicate, modify, extend, and subvert conventions, rising above the mundane works that surrounded them in their original contexts. Scale changes not only the quantities of texts that can be interrogated but also the contexts and contents of the questions.

Scale also raises questions about one of the most privileged terms in the Traditional Humanities, *reading*. At the level professional scholars perform this activity, reading is so intimately related to meaning that it connotes much more than parsing words; it implies comprehending a text and very often forming a theory about it as well. Franco Moretti (2007:56–57) throws down the gauntlet when he proposes “distant reading” as a mode by which one might begin to speak of a history of *world* literature. Literary history, he suggests, will then become “a patchwork of other people’s research, *without a single direct textual reading*” (57; emphasis in original). He continues, “Distant reading: where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems” (57). In this understanding of “reading,” interpretation and theorizing are still part of the picture, but they happen not through a direct encounter with a text but rather as a synthetic activity that takes as its raw material the “readings” of others.

If one can perform “distant reading” without perusing a single primary text, then a small step leads to Timothy Lenoir’s claim (2008a) that machine algorithms may also count as “reading.” Chapter 3 discusses machine reading in more detail, but here I note that from Lenoir’s perspective, algorithms read because they avoid what he sees as the principal trap of conventional reading, namely that assumptions already in place filter the material so that one sees only what one expects to see. Of course, algorithms formed from interpretive models may also have this deficiency, for the categories into which they parse units have already been established. This is why Lenoir proclaims, “I am totally against ontologies” (2008a). He points out that his algorithms allow convergences to become visible, without the necessity to know in advance what characterizes them.

Lenoir’s claim notwithstanding, algorithms formed from ontologies may also perform the useful function of revealing hitherto unrecognized assumptions. Willard McCarty makes this point about the models and relational

databases he used to analyze personification in Ovid's *Metamorphoses*. While the results largely coincided with his sense of how personification works, the divergences brought into view strong new questions about such fundamental terms as "theory" and "explanation" (2005:53–72). As he remarks, "A good model can be fruitful in two ways: either by fulfilling our expectations, and so strengthening its theoretical basis, or by violating them, and so bringing that basis into question" (2008:5).

The controversies around "reading" suggest it is a pivotal term because its various uses are undergirded by different philosophical commitments. At one end of the spectrum, "reading" in the Traditional Humanities connotes sophisticated interpretations achieved through long years of scholarly study and immersion in primary texts. At the other end, "reading" implies a model that backgrounds human interpretation in favor of algorithms employing a minimum of assumptions about what results will prove interesting or important.⁴ The first position assumes that human interpretation constitutes the primary starting point, the other that human interpretation misleads and should be brought in after machines have "read" the material. In the middle are algorithms that model one's understanding but nevertheless turn up a small percentage of unexpected instances, as in McCarty's example. Here human interpretation provides the starting point but may be modified by machine reading. Still another position is staked out by Moretti's way of unsettling conventional assumptions by synthesizing critical works that are themselves already synthetic (2000, 2007). Human interpretation remains primary but is nevertheless wrenched out of its customary grooves by the scale at which "distant reading" occurs. Significantly, Moretti not only brackets but actively eschews the level on which interpretation typically focuses, that is, paragraphs and sentences (2007:57).

The further one goes along the spectrum that ends with "machine reading," the more one implicitly accepts the belief that large-scale multicausal events are caused by confluences that include a multitude of forces interacting simultaneously, many of which are nonhuman. One may observe that humans are notoriously egocentric, commonly perceiving themselves and their actions as the primary movers of events. If this egocentric view were accurate, it would make sense that human interpretation should rightly be primary in analyzing how events originate and develop. If events occur at a magnitude far exceeding individual actors and far surpassing the ability of humans to absorb the relevant information, however, "machine reading" might be a first pass toward making visible patterns that human reading could then interpret.

In any case, human interpretation necessarily comes into play at some point, for humans create the programs, implement them, and interpret the results. As Eyal Amiran (2009) observes, the motors driving the process are human desire and interest, qualities foreign to machines. Nevertheless, a human interpreting machine outputs constitutes a significantly different knowledge formation than the Traditional Humanities' customary practice of an unaided human brain-body reading books and arriving at conclusions. Given that human sense-making must necessarily be part of the process, at what points and in what ways interpretation enters are consequential in determining assumptions, methods, and goals. Also at work here is the self-catalyzing dynamic of digital information. The more we use computers, the more we need the large-scale analyses they enable to cope with enormous data sets, and the more we need them, the more inclined we are to use them to make yet more data accessible and machine-readable.

That large-scale events are multicausal is scarcely news, but analysis of them as such was simply not possible until machines were developed capable of creating models, simulations, and correlations that play out (or make visible) the complex interactions dynamically creating and re-creating systems.⁵ In turn, the use of tools unsettles traditional assumptions embedded in techniques such as narrative history, a form that necessarily disciplines an unruly mass of conflicting forces and chaotic developments to linear storytelling, which in turn is deeply entwined with the development and dissemination of the codex book. As Alan Liu (2008a) aptly observes about digital technologies (equally true of print), "These are not just tools but tools that we think through." The troops march together: tools with ideas, modeling assumptions with presuppositions about the nature of events, the meaning of "reading" with the place of the human.

The unsettling implications of "machine reading" can be construed as pointing toward a posthuman mode of scholarship in which human interpretation takes a backseat to algorithmic processes. Todd Presner (2008), creator of *Hypermedia Berlin* (2006) and codirector of the *HyperCities* project, reacted strongly when I asked him if digital methods could therefore be seen as erasing the human. As he pointed out, "human" is not a fixed concept but a construction constantly under challenge and revision. Although he conceded that one might characterize certain aspects of the Digital Humanities as posthuman, he insisted the shift should be understood contextually as part of a long history of the "human" adapting to new technological possibilities and affordances. Technologically enabled transformations are nothing new, he argued. Indeed, a major theme in this book is the coevolutionary

spiral in which humans and tools are continuously modifying each other (for further elaboration, see Deacon [1998]), Stiegler [1998], Hansen [2006a], and chapters 3, 4, and 5).

The tension between algorithmic analysis and hermeneutic close reading should not be overstated. Very often the relationship is configured not so much as an opposition but as a synergistic interaction. Matthew Kirschenbaum (2009) made this point when discussing a data-mining project designed to rank the letters Emily Dickinson wrote to Susan Huntington Dickinson in terms of erotic language. In interpreting the results, Kirschenbaum and his colleagues sought to understand them by reverse-engineering the sorting process, going back to specific letters to reread them in an attempt to comprehend what kind of language gave rise to a given ranking. The reading practices consisted of what Kirschenbaum calls "rapid shuttling" (2009) between quantitative information and hermeneutic close reading. Rather than one threatening the other, the scope of each was deepened and enriched by juxtaposing it with the other.

The possibility of creating synergistically recursive interactions between close reading and quantitative analyses is also what Stephen Ramsay (2008a) has in mind when he calls for "algorithmic criticism," where the latter word implies hermeneutic interpretation. Positioning himself against a mode of inquiry that praises computer analyses for their objectivity, Ramsay argues that this "scientific" view (2008b) forsakes the rich traditions of humanistic inquiry that have developed sophisticated and nuanced appreciation for ambiguities. "Why in the world would we want the computer to settle questions?" he asks, proposing instead that computers should be used to open up new lines of inquiry and new theoretical possibilities.

Productive/Critical Theory

What might be these theoretical possibilities? Conditioned by several decades of post-structuralism, many humanistic disciplines associate "theory" with the close scrutiny of individual texts that uncovers and destabilizes the founding dichotomies generating the text's dynamics. A different kind of theory emerges when the focus shifts to the digital tools used to analyze texts and convey results. Jay David Bolter (2008) suggests the possibility of "productive theory," which he envisions as a "codified set of practices." (We may perhaps consider the work of Diane Gromola and Bolter [2003] as characteristic of productive theory.) The ideal, Bolter suggests (2008), would be an alliance (or perhaps integration) of productive theory with the