

# When Computers Read: Literary Analysis and Digital Technology

by Sarah Jones

## Digital Humanities and Information Visualization

### EDITOR'S SUMMARY

The study of literature is changing in dramatic ways, stimulated by new opportunities that digital technology presents. Data visualization upends the dynamic for literary analysis, focusing not on questions stemming from a critic's personal viewpoint but on revealing and displaying connections between elements of the literary experience. The dominant association between critic and text is downplayed, replaced with associations within the text and between it and its context. The basis of interpretation shifts from reading to seeing, from qualitative analysis to quantitative. The reader's role is transformed, as well, from following the critic's path of thinking to actively exploring a network of multisensory and interdisciplinary information. The distinction between the authoritative presenter/critic and the learner/explorer is blurred. By inviting literary scholars to ask different questions for computational analysis, digital technology and visualization inspire innovative investigations and enable new insights.

### KEYWORDS

literature  
 electronic visualization  
 quantitative analysis  
 information technology

Sarah Jones is pursuing an MA at New York University (NYU) and an MSLIS at Long Island University. She works in the NYU library. She can be reached at [scj248@nyu.edu](mailto:scj248@nyu.edu).

Digital technology has encouraged innovation not only in the methods of research but also in the presentation of results. Data visualization turns findings into charts, graphs, maps and exhibits – all forms that make literary interpretation a visually experienced object or event rather than an abstract concept. Lev Manovich, a professor of visual arts, describes the difference between traditional expository writing and data visualization: “What’s interesting about culture is that the categories are continuous. Instead of using these techniques to reduce complexity, to divide data into a few categories, I want to map the complexity” [1, p. 11].

Computational analysis promises the discipline of literature access to the kind of knowledge that was once regarded as the antithesis of the humanities: hard facts. As a result of emerging digital technologies literature can now be studied and presented with techniques that have traditionally been confined to sociology, natural science, neuroscience, history, psychology and linguistics. This paper will address two questions raised by the introduction of digital methods into the humanities: What does it mean for the literary critic? What does it mean for the reader of the resulting literary analysis?

### New Formats: From Papers to Puzzles

Imagine, for example, a web representing the conversations among characters in Hamlet. Now imagine an essay describing the structure of these connections. An essay outlining these phenomena in a linear fashion would run the risk of being either confusing or reductive. In addition, the essay's author would make the subjective personal choice of which connections to describe in further detail. Traditional literary interpretation favors a particular message, point or theme. By contrast, patterns, shapes and trends would be immediately apparent in the visual web. Data webs represent multiple angles, connections and shapes simultaneously. Literary historian

JONES, continued

Susan Brown has noted that some areas of the humanities are “rich in dense and complex interlinkages which almost defy explanation in words” [2, p.5]. In a well-crafted visual form the complexity of the data would not have to be sacrificed for the sake of clarity.

Data visualization changes the experience of literary interpretation not just for the critic but for the reader as well. It has already been noted that the traditional analytic essay is the personal commentary of a single reader who is recognized as an expert qualified to interpret texts. While the literary critic draws from a cultural knowledge and intellectual tradition, the interpretation is still very much an explanation of the personal interaction between this privileged reader and the text. Readers of the resulting criticism experience the literary work filtered through the *personal* interpretation of the theorist. While the readers must apply their own thought processes to the texts, the critic makes many decisions for them, such as which factors and connections within the literary work are worth attention. In many ways the shape of the reader’s literary experience is drawn by the critic.

A reader’s interaction with data visualization is still influenced by a privileged reader, the critic. The critic (in this situation “researcher” may be a more appropriate title) decides which questions to “ask” the computer and then designs the visual representation of the results. The outcome of the data, however, is somewhat unpredictable. The researcher determines the hypothesis and method of the experiment but cannot control the results. It is highly unlikely, for example, that Franco Moretti, a professor at Stanford University and one of the most prominent digital-humanities academics, had anticipated the results of a study he performed on literary genre: that the word *the* appeared more frequently in Gothic novels [3]. In this situation the literary critic is in the position not of documenting individual experience with the act of reading, but of presenting hard data. It is important to note that this type of analysis is still affected by the researcher’s personal bias—the subject and experiment are, after all, controlled by the critic. However, as will be discussed in greater detail, results presented in the forms offered by humanities computing have the potential to step closer to objectivity.

In reference to a study that mapped the connections among women writers in England, Brown observed that “seeing a visual representation that

summarizes a pattern in the data is different from reading the same data as a text ... an interactive visual environment is intended primarily to assist in pattern-finding” [2, p. 3]. Data visualization, then, makes the act of literary interpretation less *reading* and more *seeing*. The viewer of data visualization processes literary study differently from the reader of criticism. Brown’s usage of the word *interpret* suggests a more active role on the part of the reader/viewer. An authoritative critic may be called upon to explain data but ultimately numbers and visuals require more interpretation on the part of the reader than does expository writing. To read a critical essay is to follow the writer along a path of thought; studying a chart is a very different experience, one that belongs more to the reader.

### New Questions, New Answers

The domination of the literary canon is weakened by digital literary studies. Computers detect quantity, not quality, which means that scholars must change the nature of the questions they ask about literature. Quantitative data lends itself to inquiries like “*Why was this author commercially successful?*” rather than “*What makes this novel good?*” When studying Sir Arthur Conan Doyle, Moretti does not attempt to explain why he believes the Sherlock Holmes stories to be objectively great works of literature. Instead, he asks, “*Why did readers of the era prefer Conan Doyle’s writing to that of other detective novelists?*” He uses a “tree” to find patterns that reveal correlations between the form of a novel and its commercial success [4]. A more traditional approach would have focused on the canonized work, ignoring the less prestigious writing from which Moretti draws a great deal of his insight.

While humanities computing can allow literary studies to more closely resemble scientific inquiry, it is important to note that digital innovations also cause a structural break from both traditions. The expository essay form has much in common with the scientific method as both are structured to prove, disprove or illuminate a single point. Classifications (such as species in science and genre in literature), causal relationships and definitive conclusions are relied upon heavily in both disciplines. The essay and the experiment are linear forms that begin with a question and end with an answer.

The difference between an insight expressed in writing and one expressed

visually or numerically raises an interesting question: How do we define and structure knowledge? Is it the understanding of causal relationships or an awareness of a number of interconnected factors at play? According to media theorist Donna Haraway, traditional Western thought has sought to make sense of the world by dividing it into culturally determined categories and hierarchies [5]. We can translate this concept to the study of literature – “themes” or “messages” that critics derive from literary works function in the same way. Due to its demands for summary and conclusion, the linear essay form can be seen as both encouraging and reflecting this kind of hierarchical thought. It prompts the writer to arrange ideas vertically according to perceived importance.

In discussing the organization of human culture, Haraway recommends an alternative to the traditional stratified power structure: a network “suggesting the profusion of spaces and identities and the permeability of boundaries” [5, p.170]. If we apply this recommendation to the hierarchical form of traditional literary study, it seems that humanities computing has the potential to present the products of criticism in the form Haraway recommends. In these forms – a table of figures, a map of connections – knowledge is organized not as a line or a column, but as a tangled web.

### New Roles for Readers

New forms of literary analysis create a new kind of reader. As already discussed, digital literary scholarship takes some of the interpretative responsibility from the critic and places it on the reader. In his criticism of the written word, Plato presents the written idea as sterile. That is, it cannot grow and be shaped through interaction in the manner of an idea spoken in conversation. The idea born from interactive discussion, on the other hand, is more provocative [6]. This distinction can be applied to modern forms of scholarship, too. The expository essay speaks in a single static voice while the chart permits divergent interpretations by presenting a sort of conversation among elements in the text. This kind of insight is less about the authority telling the reader how the text functions and more about showing aspects of the work through the presentation of data. This creates a more autonomous reader who must examine a web of findings, decide which connections are significant or interesting, and then, if appropriate, conceive of possible explanations for

the pattern. While these explanations may be artificial “wholes,” the emphasis on reader-created interpretations of data leads to a number of diverse viewpoints. This makes the discipline more open to polyvocality, thereby blurring the hierarchical division between the “authority” and the “learner.”

It has already been established that the digital-humanities reader must play a more active role, but what does this responsibility mean in practice? The difference between “active” and “passive” audiences is a much-discussed subject among media theorists. Marshall McLuhan asserts that the manner in which information is presented is just as influential as the information itself. McLuhan divides formats into two categories: “hot” and “cool” media. Hot media are characterized by their tendency to demand the full attention of a single human sense and demand only minor effort on the part of the audience. Cool media contain less and more scattered data and require a more active audience who will connect the dots and fill in the blanks [7]. It is, in essence, the difference between a puzzle and a picture. Hieroglyphs (cool) and letters (hot), for example, differ from each other in that hieroglyphs are pictures representing objects. The reader must figure out the relationships among the images. Words, by contrast, are abstracted from the physical forms they represent and thus a person who understands the phonetic system is provided with all the connections necessary for a complete understanding of the message.

With McLuhan’s ideas in mind, it would seem that the switch from written theory to visual and numerical forms would mark a sea change in sense perception and have an enormous impact on the reader’s experience of literary criticism. Data visualization turns concepts into physical objects and displays them in relation to each other. Therefore, it more closely resembles a hieroglyph than it does a written word; it requires more dot-connecting and independent work on the part of the reader than would a traditional essay. The products of machine-based research are not so much *read* as they are *operated*; they are not *absorbed*, but *worked with*.

What are the implications of literary criticism that demands a more active reader? McLuhan suggests that the tendency of hot media to privilege a single sense leads to a sort of tunnel vision – a specialized and fragmented worldview [7]. This state of affairs bears a striking resemblance to traditional humanities studies in which literature is isolated from other

JONES, continued

disciplines and their corresponding senses. With this in mind it makes perfect sense that the cooling-down of literary studies would coincide with an increased interest in work that is interdisciplinary, multisensory and large in scope. When critics and readers must connect numerical data to historical context and literary texts, there is an increased need for exploration in areas such as science, statistics, linguistics and history.

### Conclusion: What's Next for Critics and Readers?

Both readers and critics must now interact with literary insight not just through the written word, but as a multisensory and interdisciplinary experience. The digital-humanities readers are active because they must connect the dots between pieces of data; when working with the new media forms produced by the digital humanities, critics, too, must adopt new methods of making connections. Hard data lends itself to questions that are more about the grand scope – the cultural and historical – and less about the personal – the moral and emotional – effect of a text. The digital-humanities critic, then, must focus on connections between texts and texts and between texts and culture; the connection between the critic and the text is deemphasized.

For the critic (or “writer/researcher/designer”), humanities computing encourages movement that breaches the confines that characterize close reading, including personal viewpoint, time constraints, historical situation and disciplinary subject boundaries. While no theory can completely escape the influence of the personal and cultural context of the critic, digital and

quantitative methods allow literary analysis to shed some of its subjectivity and move along the continuum towards objectivity. For the reader (or “viewer/player/listener”) of literary analysis, humanities computing can have the opposite effect. The reader must take an active role in assessing the meaning of the data when the results of literary analysis are presented in a form that does not contain a prepackaged conclusion. The products of digital literary study encourage the reader to make a number of individual choices. As a result, the reader’s experience of literary criticism is actually more personal in a computational context.

For both the critic and the reader, the digital humanities provide a new conception of the world of literature. Not only is this world larger – the sheer volume of the material we can access is unprecedented – but it is open to levels of analysis that could never be achieved by human brainpower alone. Hierarchies and themes fade into the background as patterns and networks emerge. These methods simultaneously divide texts into new categories and connect them to each other to form new wholes. As digital innovations progress, literary scholars and their audiences must work through new issues emerging in their discipline. What should critics do when humanities computing produces inexplicable results? How will active readers change the field of literary interpretation? What tools and skills should the critic and the reader acquire to facilitate their interactions with charts and statistics? In the face of hard data, the most important question is this: What kinds of insight are valuable to the study of literature? ■

### Resources Mentioned in the Article

- [1] Williford, J. (2011). Graphing culture. *Humanities*, 32(2), 11-53. Retrieved February 14, 2012, from [www.neh.gov/news/humanities/2011-03/graphing.html](http://www.neh.gov/news/humanities/2011-03/graphing.html).
- [2] Brown, S. (2009). Visualizing varieties of association in Orlando. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(1), 1-5. Retrieved February 14, 2012, from <https://letterpress.uchicago.edu/index.php/jdhcs/article/view/7>.
- [3] Schulz, K. (June 24, 2011). What is distant reading? *The New York Times*. Retrieved February 14, 2012, from [www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html](http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html).
- [4] Moretti, F. (March 2000). The slaughterhouse of literature. *Modern Language Quarterly*, 61(1), 207-27.
- [5] Haraway, D. (1991). *Simians, cyborgs, and women: The reinvention of nature*. New York, NY: Routledge.
- [6] Plato. (2002). *The Phaedrus*. (R. Waterfield, Trans.). New York, NY: Oxford University Press.
- [7] McLuhan, M. (1994). *Understanding media: The extensions of man*. Cambridge, MA: The MIT Press.