

BOOKS | HUMANITIES 2.0

Analyzing Literature by Words and Numbers

By PATRICIA COHEN DEC. 3, 2010

Victorians were enamored of the new science of statistics, so it seems fitting that these pioneering data hounds are now the subject of an unusual experiment in statistical analysis. The titles of every British book published in English in and around the 19th century — 1,681,161, to be exact — are being electronically scoured for key words and phrases that might offer fresh insight into the minds of the Victorians.

This research, which has only recently become possible, thanks to a new generation of powerful digital tools and databases, represents one of the many ways that technology is transforming the study of literature, philosophy and other humanistic fields that haven't necessarily embraced large-scale quantitative analysis.

Dan Cohen and Fred Gibbs, the two historians of science at George Mason University who have created the project, have so far charted how frequently more than two dozen words — among them “God,” “love,” “work,” “science” and “industrial” — appear in British book titles from the French Revolution in 1789 to the beginning of World War I in 1914. To Mr. Cohen, the sharply jagged lines that dance across his graphs can be used to test some of the most deeply entrenched beliefs about the Victorians, like their faith in progress and science: “We can

Access more of The Times by creating a free account or logging in.

[CONTINUE](#)[Subscriber Login](#)

Mr. Cohen said that he and Mr. Gibbs hoped that their work could serve as a model for how scholars might use the shopping cart of new digital tools to challenge longstanding assumptions and interpretations across the humanities.

Some of their colleagues are clearly intrigued by the possibilities.

“My own reaction was sheer exhilaration,” said Alice Jenkins, a professor of Victorian literature and culture at the University of Glasgow, who saw Mr. Cohen present his preliminary results at a recent conference on the Victorians.

There is also anxiety, however, about the potential of electronic tools to reduce literature and history to a series of numbers, squeezing out important subjects that cannot be easily quantified.

“I was excited and terrified,” said Matthew Bevis, a lecturer at the University of York in Britain, who was at the same conference. “This is not just a tool; this is actually shaping the kind of questions someone in literature might even ask.”

“It should come in a box marked ‘Handle With Care,’ ” he added.

Such concerns didn’t stop Mr. Bevis or other academics in the audience from asking Mr. Cohen to run a few electronic searches of particular words pertinent to their own work. Meredith Martin, an assistant professor of English at Princeton who is studying the history of poetic form, was interested in the terms “prosody,” “meter” and “verse.”

“I actually sent him an e-mail as he was talking,” Ms. Martin said. She figured he would be inundated with requests, and “I wanted to be first in line.”

Mr. Cohen and Mr. Gibbs’s “Reframing the Victorians” study is one of 12 university projects to win a new digital humanities award created by Google that provides money along with access to the company’s powerful computers and databases.

Some scholars are wary of the control an enterprise like Google can exert over digital information. Google’s plan to create a voluminous online library and store has raised alarms about a potential monopoly over digital books and the hefty pricing that might follow.

Access more of The Times by creating a free account or logging in.

CONTINUE

[Subscriber Login](#)

But Jon Orwant, the engineering manager for Google Books, Magazines and Patents, said the plan was to make collections and searching tools available to libraries and scholars free. “That’s something we absolutely will do, and no, it’s not going to cost anything,” he said.

One criterion in choosing projects to finance, he added, was whether they were going to create new data sets and computer codes that other researchers would find useful.

Mr. Gibbs and Mr. Cohen’s searches of book titles represent only an initial swipe at the data. Step 2 is canvassing the full texts. The professors will also have the ability to zero in on details, specific titles and passages.

Their starting point was an earlier work that focused on the written word as an entry point into the era: Walter E. Houghton’s “Victorian Frame of Mind, 1830-1870,” a landmark book published in 1957 that has shaped generations of scholarship, even as its conclusions have been challenged. Mr. Houghton sought to capture what he called a “general sense” of how middle- and upper-class Victorians thought, partly by closely reading scores of texts written during the era and methodically counting how many times certain words appeared. The increasing use of “hope,” “light” and “sunlight,” for instance, was interpreted as a sign of the Victorians’ increasing optimism.

Mr. Houghton’s reading list was monumental, yet his methodology raised questions about the validity of extrapolating the attitudes of millions of people from a couple of hundred texts.

The kind of comprehensiveness that digital research offers quells such complaints. “All history is anecdotal,” Mr. Cohen said. “You could read three books and say the Victorians were really obsessed with evil, or you could read 30 books, or 300 books; but you didn’t read 10,000 books.”

But now, he explained, vast digital libraries present “for the first time the possibility that we can conduct a comprehensive survey of Victorian writing — not just the well-known Mills and Carlyles, but tens of thousands of lesser-known or even forgotten authors.”

Access more of The Times by creating a free account or logging in.

CONTINUE

[Subscriber Login](#)

and “universal” is consonant with the conventional view that the 19th century was a time of rising secularism and skepticism.

Yet large searches can also challenge some pet theories of close reading, he said: for example, that the Victorians were obsessed with the nature and origins of evil. As it turns out, books with the word “evil” in the title bumped along near the bottom of the graph, accounting for less than 0.1 percent — a thousandth — of those published during the Victorian era.

As Mr. Cohen is quick to acknowledge, the meaning of those numbers is anything but clear. Perhaps authors didn’t like to use the word “evil” in the title; perhaps there were other, more common synonyms; perhaps the context points to another subject altogether.

Ms. Martin at Princeton knows firsthand how electronic searches can unearth both obscure texts and dead ends. She has spent the last 10 years compiling a list of books, newspaper and journal articles about the technical aspects of poetry.

She recalled finding a sudden explosion of the words “syntax” and “prosody” in 1832, suggesting a spirited debate about poetic structure. But it turned out that Dr. Syntax and Prosody were the names of two racehorses.

“You find 200 titles with ‘Syntax,’ and you think there must be a big grammar debate that year,” Ms. Martin said, “but it was just that Syntax was winning.”

Scholars should also remember that the past contains more than the written record, Mr. Bevis said in an interview. Fewer references to a subject do not necessarily mean that it has disappeared from the culture, but rather that it has become such a part of the fabric of life that it no longer arouses discussion. He quoted Emily Dickinson: “Is it oblivion or absorption when things pass from our mind?”

Of more concern, Mr. Bevis said, is the fear that statistical measures could overshadow meaning and interpretation.

Access more of The Times by creating a free account or logging in.

CONTINUE

[Subscriber Login](#)

scale, quantitative research is likely to highlight “the importance and the value of close reading; the detailed, imaginative, heightened engagement with words, paragraphs and lines of verse.

“Close reading,” she continued, “will become even more crucial in a world in which we can, potentially, read every word of Victorian writing ever published.”

Humanities 2.0: This is the second article in a series about how digital tools are changing scholarship in history, literature and the arts. The first article examined projects that mine vast stores of digitized materials and the debates over the usefulness of such research.

A version of this article appears in print on December 4, 2010, on Page C1 of the New York edition with the headline: HUMANITIES 2.0; Analyzing Literature by Words, but Also by the Numbers.

© 2019 The New York Times Company

Access more of The Times by creating a free account or logging in.

CONTINUE

[Subscriber Login](#)