Professor uses big data to research history of gender in fiction

February 28 2018

The number of women writing works of fiction dropped dramatically from the middle of the 19th century to the middle of the 20th century, and the prominence of female characters in works of fiction declined as well.

At the same time, however, the gender differences between male and female characters became weaker. Ted Underwood, a University of Illinois professor of information sciences and of English, came to those seemingly conflicting findings when he used data-mining tools to look at 104,000 books written over a period of more than 200 years.

Underwood and his colleagues, David Bamman of the University of California, Berkeley and U. of I. graduate student Sabrina Lee, explored the significance of gender in fiction by using an algorithm to look at books in the HathiTrust Digital Library. Their findings are published in the Journal of Cultural Analytics.

Looking at how much space was devoted to male and female characters, the researchers saw a steady decline in the space devoted to women from 1800 to 1960, "in the very period when we might expect to see the effects of first-wave feminism."

At the same time, they wrote, women authors were losing shelf space. They found a "fairly stunning decline in the proportion of fiction writers who were women," from about half of all fiction books being written by women in 1850 to barely a quarter in 1950.

One theory for the decline is that fiction writing was dominated by women in the early 19th century when it was not a high-status career. As the prestige of the novelist increased, more men moved into writing fiction. At the same time, the researchers write, more intellectual opportunities other than novelist were becoming available for women.

Male authors devote less space in their novels to female characters, who account for one-quarter to one-third of the character space, the research showed. The division of space devoted to male and female characters is nearly equal in novels written by women.

"Men write stories where there are not that many women. Women represent the world as it is, with equal numbers of men and women, and men just don't," Underwood said.

"We see no progress over 200 years in the overall number of characters in fiction who are women, even with multiple waves of feminism and social change. Victorian literature is every bit as balanced as our world" in terms of the number of female characters and the space devoted to discussing them, he said.

But a greater proportion of male authors doesn't account for all the underrepresentation of women in fiction, Underwood said. When he and his colleagues looked at female characters in novels written by women, they found those characters were becoming somewhat less prominent even in books written by women.

Underwood said the increase in genre fiction – Westerns and adventure stories, for example – may play a part in the trend toward less space for female characters.

The way male and female characters are represented in fiction has become less sharply drawn from the mid-19th century to today, though.

The researchers looked at the adjectives used to describe characters and the verbs that described their actions. In the 19th century, the language of thought and feeling was feminine. Women characters "felt" and were described by words such as heart and spirit, while men more often "got." Women were associated with private spaces such as chambers and rooms, while men were associated with houses and countries.

Male authors tend to portray gender differences more clearly than female authors. "Gender stereotypes are decreasing in male fiction too, but women lead the way," Underwood said.

Although gender differences became increasingly blurred, there are still certain descriptions that are strongly gendered, he wrote. In a mid-20th-century quirk of language, women smiled and laughed in stories, while men only grinned and chuckled, and their grins were often menacing. In physical descriptions, references to hair are nearly always female, while 20th-century male characters have pockets they are constantly putting things in.

Underwood would not be able to ask large-scale questions about literary history over a broad timeline without machine learning and access to a large digital library.

"Machine learning allows us to pose questions about concepts, like gender, that lack a clear definition," he said. "Models using evidence from different historical periods can learn to define masculinity or femininity differently.

"The HathiTrust Digital Library is a great resource. We wouldn't have been able to say anything much after 1923 without HathiTrust sharing information from those volumes, because they are under copyright."

The researchers have shared the dataset they used and Underwood hopes

others will use it to pose new questions about the history of gender in fiction.

Provided by University of Illinois at Urbana-Champaign