Women in novels have tended to “feel”, while men “get”; women smile or laugh, while men grin or chuckle. An analysis of more than 100,000 novels spanning more than 200 years shows how gendered even seemingly innocuous words can be – as well as revealing an unexpected decline in the proportion of female novelists from the middle of the 19th century to the middle of the 20th.

Academics from the University of Illinois and the University of California, Berkeley used an algorithm to examine 104,000 works of fiction from 1780 to 2007, drawn mostly from HathiTrust Digital Library. The algorithm identified both authors’ and characters’ genders.

The academics expected to see an increase in the prominence of female characters in literature across the two centuries. Instead, “from the 19th century through [to] the early 1960s we see a story of steady decline”, said Ted Underwood, David Bamman, and Sabrina Lee in their research paper, The Transformation of Gender in English Language Fiction, which has just been published in Cultural Analytics.

As well as the drop in the number of characters who are women or girls, they also found “a fairly stunning decline” in the number of books written by women in the first half of the 20th century, writing that “the proportion of fiction actually written by women ... drops by half (from roughly 50% of titles to roughly 25%) as we move from 1850 to 1950.”

The academics were so surprised by these findings – “in the very period when we might expect to see the effects of first-wave feminism” – that they thought it must be down to an error in their methods. They ran further tests, and found they tallied.
“Women go from representing almost half the authors of fiction to barely a quarter. If this trend is real, it is an important fact about literary history that ought to be foregrounded even, say, in anthology introductions.”

The academics speculate that one reason for the drop, which reverses in about 1970, could be the “gentrification” of the novel. In the mid-19th century, novel-writing was not a “highstatus career”, but as it increasingly became so, it became more desirable to male writers.

Kate Mosse, a bestselling historical novelist and founder of the Women’s Prize for fiction, said she was not surprised by the results. She pointed to “a sea change from the Enlightenment through to Victorian values, so women are freer in the time of Jane Austen or Mary Shelley or Ann Radcliffe but then Victorian values, the idea of the angel in the home, take over.”

While representation of women in fiction fell, non-fiction saw “an enormous expansion” of female writers. “Expanding opportunities on this scale might have lured women away from the novel,” they write.

The decline in women writing is part of the reason for the drop in female characters, the analysis showed. In books by men, women occupy on average of just a quarter to a third of the character space. “The division is much closer to equal” in books by women.

“This gap between the genders is depressingly stable across 200 years.”

Analysing the HathiTrust archive’s 104,000 novels, and looking at characters’ vocabulary, the researchers found that in early 19th-century novels, words such as “heart”, “tears”, “sighs”, “smiles” and “spirits”, were “gendered feminine”, with “only a few subjective nouns ascribed more often to men; the primary one is passion, which is sometimes a 19th-century euphemism for lust”.

Fictional men have historically been more likely to have “got” things, and women more likely to have “felt”. By the middle of the 20th century, words for mirth such as “smile” and “laugh” were more likely to be applied to female characters, while “midcentury men, apparently, can only grin and chuckle”.

“This gendering of mirth peaks in the years before and after world war two, and Raymond Chandler is a typical expression of its consequences for men. His male characters have a habit of grinning in an uneasy laconic way,” the academics write.

Overall, the researchers found that gender divisions between characters had “become less sharply marked” over the last two centuries. They reiterate, however, that this blurring of the boundaries of gender as fiction moves into the 20th century “doesn’t seem to have been associated with greater emphasis on women as characters. On the contrary, their prominence declines across the same period.”

And “men remain – on average, as a group – remarkably resistant to giving women more than a third of the character-space in their stories.”