Title	Computing differences in language between male and female authors
Authors	O'Sullivan, James
Publication date	2017-10-19
Original Citation	O'Sullivan, James (2017) 'Computing differences in language between male and female authors', RTÉ Brainstorm, 19 October.
Type of publication	Contribution to newspaper/magazine
Link to publisher's version	https://www.rte.ie/eile/brainstorm/2017/1018/913307-computing-differences-in-language-between-male-and-female-authors/
Rights	© 2017 RTÉ Brainstorm; James O'Sullivan. This article is made available under a "No Derivatives" Creative Commons license - https://creativecommons.org/licenses/by-nc-nd/3.0/ie/
Download date	2025-09-11 11:58:17
Item downloaded from	https://hdl.handle.net/10468/4929

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Computing differences in language between male and female authors

James O'Sullivan

6-7 minutes

Updated / Thursday, 19 Oct 2017 **11:51**

"Many of the words favoured by our female authors are family-oriented, but they also favour interactive terms"

It's a fact that female authors still battle with stereotypes. The gender bias revealed in [an analysis of some 10,000 book reviews by Andrew Piper and Richard Jean So](#) is sufficient evidence of such assertions. As far as reviewers for [The New York Times](#) and [Sunday Book Review](#) are concerned, "women write about family, men write about war".

The results of this study are damning. Reviewers favour terms like "husband", "marriage", and "beauty" when describing female books, "theory" and "argument" when critiquing men. Piper and So conclude that 19th century attitudes have endured, that women "are still being defined by their 'sentimental' traits and a love of writing about "maternal" issues, while men are most often being defined by their attention to matters of science and the state".

Are these findings reflective of enduring social predispositions around representations of gender, or do they emerge out of visible trends across literary style? Are reviewers writing differently about men and women because men and women write differently?

> Modernist women showed the greatest variance in word choice, suggesting that female authors were the greatest contributors to the rampant experimentation at its height

This very proposition is dangerous in itself because the question "do men and women write differently?" is itself based on an assumption of difference. Nonetheless, it is a question which scholars have been asking for some time and, in recent years, there has been a rising interest in this topic amongst stylometrists, literary scholars who examine style using computational methods.

In 2008, David L. Hoover, one of the stylometry's pioneers, used what he calls Craig Zeta (named for its inventor, Hugh Craig) to compare male and female poets. A Zeta analysis produces lists of words distinct to a group of texts, relative to the comparative set. In other words, Hoover's study produced a list of words favoured by male authors, and typically
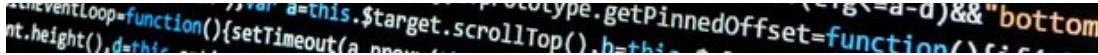
avoided by female writers, as well as those words favoured by female authors that were avoided by males.

In his findings, Hoover points to the "almost stereotypical" words used by both groups: female authors write about "children" and "mirrors", their male counterparts about "beer" and "lust". Hoover's study is largely illustrative, a fragment within a broader essay designed to outline the potential of such methods. A comprehensive dedicated study was later completed by Jan Rybicki, who traced authorial styles from the 18th to the 21st century which demonstrated that gender signals have become less discernible.

In a Humanities where computer-assisted methods are becoming increasingly prominent, the role of the human has never been more important

My colleague, Sean G. Weidman, suggested that we use a set of 236 novel-length texts written by 54 authors to reassess and expand upon these previous studies. Works were drawn from three literary epochs: Victorian, modern and contemporary. In many respects, our Zeta analyses support Hoover's findings and the resulting wordlists reveal thematic consistencies that one might consider "stereotypical".

```
theventLoop=function(){setTimeout(a   var a=this.$target.scrollTop().b=thi   getPinnedOffset=function()&& "bottom
t.height(),d=thi
```

However, there are nuances to be considered. Piper and So show that book reviewers still see women as writing about "family" and that they "obsess over love of themselves ('me')". Yes, many of the words favoured by our female authors are family-oriented, but they also favour interactive terms, more selfless language that we consider holds more of an external focus.

Our study also demonstrates that pronoun usage amongst female authors increases over time. The use of "me" and "I" are not indicative of an "obsession with themselves", but represent post-Victorian efforts by women writers to invigorate the feminist project through literature. It is also an effort to carve out a space for under-represented voices and stories and marginalised styles.

Are reviewers writing differently about men and women because men and women write differently?

Interestingly, modernist women showed the greatest variance in word choice, suggesting that female authors were the greatest contributors to the rampant experimentation of the epoch at its height. One could view these results as the quantification of literary feminism, which found its voice during the modern era, before emerging as a stronger stylistic force within contemporary writing.

But studies like these beg the question: what can computers really tell us about gendered language? As a method, Zeta is an inherently dichotomous technique – that is, it is designed to produce variance – and it will always detect differences between two sets of texts.

Consequentially, the significance of findings depends greatly on the ways in which the test sets are constructed. Using it to compare male and female authors transposes dangerous social oppositions with statistically-valid results.

Furthermore, computer-assisted methods of this sort omit much context. Women may well be writing about "children" and "mirrors", but in what respect? As with any of act of distant reading, a set of distinct words can be readily misinterpreted. The responsibility is on the critic to ensure the legitimacy of subjective comparisons and literary interpretations. In a Humanities where computer-assisted methods are becoming increasingly prominent, the role of the human has never been more important.