

NATURE

Wikipedia shapes language in science papers

Experiment traces how online encyclopaedia influences research write-ups.

Mark Zastrow

26 September 2017

Wikipedia is one of the world's most popular websites, but scientists rarely cite it in their papers. Despite this, the online encyclopedia seems to be shaping the language that researchers use in papers, according to an experiment showing that words and phrases in recently published Wikipedia articles subsequently appeared more frequently in scientific papers.

Neil Thompson, an innovation scholar at the Massachusetts Institute of Technology in Cambridge, says that this finding runs counter to an academic culture that downplays Wikipedia's credibility as a knowledge source. "Academia is fighting Wikipedia," he says. Many universities, including his own, warn students against citing the website as a source in assignments. But the study, posted on the Social Science Research Network (SSRN) preprint server on 20 September and which Thompson co-authored, shows how Wiki articles can serve as constantly updated open access review articles. "In its best form, that's what Wikipedia could be," says Thompson.


Thompson and co-author Douglas Hanley, an economist at the University of Pittsburgh in Pennsylvania, commissioned PhD students to write 43 chemistry articles on topics that weren't yet on Wikipedia. In January 2015, they published a randomized set of half of the articles to the site. The other half, which served as control articles, weren't uploaded.

By February 2017, the chemistry articles had together received more than 2 million views. The researchers then analysed the text of 50 of the highest-impact chemistry journals published by Elsevier to see whether the language used in scientific papers had shifted by November 2016, nearly two years after the Wikipedia articles were posted.

Using text-mining techniques to measure the frequency of words, they found that the language in the scientific papers drifted over the study period as new terms were introduced into the field. This natural drift equated to roughly one new term for every 250 words, Thompson told *Nature*. On top of those natural changes in language over time, the authors found that, on average, another 1 in every 300 words in a scientific paper was influenced by language in the Wikipedia article.

The influence of Wikipedia was more apparent in less-cited journals than in the most well-known publications. The authors suggest that ideas and language first published on topics entirely new to science make their way into Wikipedia before feeding back into the literature in follow-up studies, published in less-frequently cited journals. When the authors analysed papers by the author's country, they found the effect was greater in lower-income countries compared to higher-income countries. Hanley says some authors may be more reliant on Wikipedia if they have limited access to expensive journals. In this way, Wikipedia serves as an equaliser, extending science to those with less resources, he says.

Adam Dunn, a data scientist at Macquarie University in Sydney, Australia, calls the study's randomized controlled trial an "ingenious" idea. But he questions the authors' claim that Wikipedia is shaping the ideas of science. **He thinks the study shows that scientists refer to Wikipedia as a way of standardizing their language when they write papers.** "It probably is showing an effect of Wikipedia, but I'm not sure the claim is what they're suggesting," he says.



Perhaps scientists, like students, are also using Wikipedia to write their scholarly articles!