It's not just you: science papers are getting harder to read

Papers from 2015 are a tougher read than some from the nineteenth century — and the problem isn't just about words, says Philip Ball.

30 March 2017

Philip Ball

Modern scientific texts are more impenetrable than they were over a century ago, suggests a team of researchers in Sweden. It's easy to believe that.

You can be confident, for example, that if you pick up a random copy of *Nature* (which has long prided itself on the relative accessibility of its papers), you may find sentences like this in the abstracts:

Here we show that in mice DND1 binds a UU(A/U) trinucleotide motif predominantly in the 3' untranslated regions of mRNA, and destabilizes target mRNAs through direct recruitment of the CCR4-NOT deadenylase complex.

But this type of jargon-heavy phrasing is not the only problem that neuroscientist William Hedley Thompson and his colleagues at the Karolinska Institute in Stockholm are tackling in their analysis. They scoured more than 700,000 English-language abstracts published between 1881 and 2015 in 122 leading biomedical journals. Their study[1], posted on the preprint server bioRxiv on 28 March, suggests that it's not just the technical jargon that has been on the rise.

There has also, the authors say, been an increase in "general scientific jargon": that is, multisyllable words that have non-technical meanings but have become part of the standard lexicon of the science paper. These words include 'robust', 'significant', 'furthermore' and 'underlying' — all familiar enough in daily use, but markedly more prevalent in the scientific literature. The words aren't inherently opaque, but their accumulation adds to the mental effort involved in reading the text.

Fourth-grade readers

Thompson and his colleagues examined the texts using standard indicators of reading ease, which measure factors such as the number of syllables per word, the number of words in a sentence and the number of words in the paper not included in a predefined list of common words (the New Dale–Chall, or NDC, list). By these measures, the trends seem very clear: a steady and marked decline in readability since 1881.

You could argue over the technicalities of the study. The list of common words is measured against the comprehension skills of US fourth-graders — children aged nine and ten — and its applicability to the scientific literature is not clear, points out Yellowlees Douglas at the University of Florida in Gainesville, author of *The Reader's Brain* (Cambridge University Press, 2015), a writing manual that bases its advice on neuroscience. More problematically, metrics such as syllable-counting are too simplistic: for example, they rate words such as 'orange' and 'praxis' as equivalent.

What's more, says physicist Luís Amaral of Northwestern University in Evanston, Illinois, who has studied trends in scientific publications, the data from before 1960 are too sparse and variable to be very reliable, and the trends in the data from after 1960 are less clear.

Besides, Amaral says, distinguishing between technical and general jargon isn't easy. And an increase in technical jargon is not wholly a bad thing: it can reflect the appearance of useful concepts and techniques in science. 'Sequencing' was probably not part of the scientific vocabulary before 1980, he says, but who would complain about its presence now?

However, the need for specialized terms cannot completely explain the increased impenetrability of modern scientific literature.

Thompson and his colleagues are probably right to pin some of that opacity on a habitual, almost ritualistic use of 'power words' such as 'distinct' and 'novel'. But a focus on word-counting risks distracting from what really matters about good writing.

Tangled sentences

Short, common words can be used to write sentences that are awfully hard to understand, simply because of poor grammatical construction. This is the point of Douglas's book: the reader expects to encounter concepts in a particular order, without having to search back for the right noun to go with a verb, or having to untangle intervening information-filled clauses. You can always look up jargon, but with a poorly constructed sentence you're on your own.

And comprehensibility isn't just about what a paper says, but also about what it leaves out. As a regular reader of research papers, I am often staggered by their leaps of reasoning or omission of key details, especially when I discover that these gaps are no less real to experts.

So how could the readability of scientific papers be improved? First, by recognizing that good writing doesn't happen by magic. It can be taught — but rarely is. Douglas suspects that many first drafts of papers are written by junior members of a research team who, lacking any model for what good writing looks like, take their lead from what is already in the journals. And there "they see the jargon and complexity as markers of what passes as scientific writing", she adds. Such self-reinforcing mimicry could certainly account for the trends highlighted by Thompson and his colleagues.

So where do you find good models of writing? Obviously, from good writers — not necessarily in the sciences, but anywhere[2]. There is hard evidence that sophisticated readers make sophisticated writers[3]. Why not encourage students to put down *Nature* and pick up Darwin, Dawkins or Dickens?

Nature | doi:10.1038/nature.2017.21751

References

1. Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. Preprint on bioRxiv at http://doi.org/10.1101/119370 (2017).

2. Gee, H. *Nature* **431**, 411 (2004).

3. Douglas, Y. & Miller, S. *Int. J. Bus. Admin.* **7**, 71–80 (2016).

SPRINGER NATURE