Clear as mud

It’s not surprising that some academic papers seem to swim before our eyes — the scientific literature has become steadily less accessible over the past half-century. Can we stop this trend, asks Jonathan Knight.

"There is no form of prose more difficult to understand and more tedious to read than the average scientific paper," wrote Francis Crick in his 1994 book *The Astonishing Hypothesis*. The observation is a caution to lay readers tempted to delve into the papers referenced in the book. But the co-discoverer of the structure of DNA was also acknowledging what everyone in science knows: research papers can be a nightmare to read.

It wasn’t always so. Crick and others of his generation, who began writing scientific papers in the 1940s, have witnessed the transformation of scientific prose. A form that was as readable as the average newspaper has, in some fields, become a jargon that even those familiar with the territory struggle to understand.

The balkanization of science into sub-disciplines, each with its own vocabulary, is largely to blame. Many journals are trying to tackle this, producing easy-to-read summaries of papers, and linking online papers to web-based glossaries. But these approaches tend to have a limited impact, unless addressing other factors — notably writing style — could transform many papers. Writing takes practice, yet it is not part of standard scientific training. So could science become readable again if researchers went back to school and took writing lessons?

Readability itself is not easy to quantify. Microsoft’s Word program features the Flesch Reading Ease scale, which measures the average length of words and sentences to calculate the number of years of education needed to comprehend a document. But such tools fail on several counts. For one, a long sentence that walks the reader down a path to its conclusion can be easier to follow than a muddled short sentence. And common words can be relatively long — technological or professor, for example — whereas many technical terms are short, such as meson, genome or glycan.

The common touch

Language experts generally agree that a better measure of accessibility is whether a piece of writing contains words in common usage — those that are at the front of the reader’s mind, rather than tucked away in the recesses of memory. As a general principle, the greater the percentage of common words an article contains, the easier it is to comprehend.

Donald Hayes, an emeritus professor of sociology at Cornell University in Ithaca, New York, has used this principle for more than 20 years to analyse texts. He calls it lexical difficulty, and has developed a numerical scale, known as LEX, to quantify it. The scale is based on the *American Heritage Word Frequency Book*, which ranks 87,000 words by their frequency of use in textbooks, novels, magazines and encyclopaedias from US grammar schools in 1969.

Although the ranking is over 30 years old, it remains the primary word-frequency reference. The 100th most common word, with 10,000th place. Among the scientific terms common enough to be included are ‘bacteria’ at 3,564, near ‘pump’ and ‘toll’, and ‘neuron’, which ranks 23,595 — about as common as ‘diddle’.

When calculating LEX scores, Hayes ignores the first 75 most common words as these contain little useful information. He then plots the ‘cumulative proportion’ of each word against the log of its rank. The cumulative proportion of, say, the 100th most common word — ‘know’ — is the percentage of the text made up of the words that lie between 75 and 100 in the frequency ranking. The graph (right) shows that the 100 most common words make up about 70% of all the words used by mothers when speaking to their children (orange line). In contrast, the same words make up only 20% of those used in the average Nature research article (blue line).
LEX values are generated by comparing the text’s curve with the benchmark curve for newspapers, which have a LEX score of 0. The area under the text’s curve is subtracted from the area under the newspaper curve to give the LEX value. Texts that use common words more frequently than newspapers have curves that rise rapidly, giving them a large area and a negative value; those that are skewed towards rare words end up with a positive value.

In a 1992 analysis, Hayes found that fiction for nine-year-olds scored about −32, and a transcript of farm workers talking to dairy cows — “Let’s go. Over here. You dummy, over here” — had a value of −59. Scientific papers in Nature and Science scored about 30. When Nature asked Hayes to repeat his analysis last year, papers in both journals had risen to the mid-90s. This trend is not new: in the early 1990s, papers in Science and Nature had accessibility scores of close to zero (see graph, above right), similar to those of newspapers such as The Daily Telegraph and The New York Times.

Alphabet soup

What happened, says Science’s editor-in-chief Donald Kennedy, is that sometime after the Second World War, the number of people active in science increased dramatically, creating new subdisciplines. As they entered ever more specialized fields, new vocabularies arose. The subdisciplines of biology are among the worst for jargon. In the past 20 years, immunologists have uncovered a new world of proteins and processes, each requiring a new name or acronym. Cell-signalling research is also packed with unfamiliar terms. The average paper in Cell, for example, has a LEX score of about 40.

The physical sciences do a bit better. Earth scientists often use relatively common words to describe what they study, such as ‘keel’ or ‘volcano’. Specialized vocabulary exists, but there is less of it. And according to a recent unpublished study by Hayes, average papers in Physical Review D and the astronomy journal Icarus have LEX values of about 22.

The effects of an increasingly opaque literature are easy to imagine, if difficult to quantify. If opening paragraphs or abstracts are difficult to understand, researchers may miss opportunities for collaboration between disciplines. If whole papers are unclear, students get diverted to other interests and the public’s fear and mistrust of science, which in part arises from difficulties in understanding new research, may increase.

Some journals are taking small steps to tackle the problem. Earlier this year, Science began adding one-line explanations of its papers to its table of contents. Development and the Journal of Cell Science, together with other journals published by the Company of Biologists in Cambridge, UK, have added a section to highlight half-a-dozen papers in each issue in language that is accessible to all biologists. Nature and Science have similar sections, which are complemented by longer pieces written by other academics discussing the newly published papers.

The Internet is also playing an important part in the solution. Each week, one of Science’s ‘Perspectives’ — a commentary on a published paper — in its online edition appears with links from technical terms in the text to web glossaries or sites with further information, a practice also followed by the review journals of the Nature Publishing Group. Articles in the forthcoming online journals of the Public Library of Science, a San Francisco-based organization that promotes free access to scientific literature, will be paired with lay-language summaries. And Cell Press journals now include general-interest summaries of articles in tables of contents sent out by e-mail.

But there are not perfect solutions. Scientists can be suspicious of lay summaries, fearing that they are oversimplified or inaccurate. And the Internet could exacerbate, instead of lessen, the balkanization within science. Scientists reading online are less likely to scan eye-catching figures or cogent abstracts that might entice them out of their field. And although summaries and web links lower the barrier to understanding caused by jargon and acronyms, they can’t eliminate it. “These are Band-Aids,” says Kennedy. “It will be tremendously difficult to solve the problem.”

It is also easy to forget why jargon is there in the first place. Technical terms are problematic for outsiders, but they are indispensable for specialists. They allow accurate shorthand for substances and processes that would take paragraphs to define. Apart from adding brief notes of explanation where space permits, the editors of top journals contacted for this article all agreed that there is little that can be done about jargon in research papers — it is here to stay.

Jargon busters

But there are other ways to improve readability. "Jargon is less pernicious if you can understand what is going on," says writing instructor Judith Swan of Princeton University in New Jersey, a former biochemist who now runs workshops to help scientists to improve their papers. Swan’s courses stem from a collaboration with George Gopen, a lecturer in English at Duke University in Durham, North Carolina, in which the pair developed principles of clear scientific writing by analysing published papers.

Researchers who attend the workshops expect to be told never to use the passive voice or split an infinitive, but Swan takes a different approach. “The one rule I subscribe to is that there are no rules,” she says. “One doesn’t follow rules, one exercises judgement.”

Take passive voice. Active sentences do pack more punch, says Swan, but passive ones are sometimes clearer. For example, there is no need to begin every sentence with “We.” Scientific papers tell stories about experiments and data, not scientists. Rather than use “We found the value to be x”, suggests Swan, “The value was found to be x”, suggests Swan. “The passive is a marvellous way to hide agency when agency is not important,” she says.

In general, Swan and Gopen recommend...
focusing attention on the expectations of the readers. Linguists know that information is easier to interpret if it is placed where readers expect it to be. So, for example, when a subject is introduced in a sentence, readers expect to find a verb soon after it. Everything that comes between the subject and the verb gets little attention.

A place for everything
Take this sentence from a paper in a recent issue of Science: “The emergence of virulent Plasmodium falciparum in Africa within the past 6,000 years as a result of a cascade of changes in human behaviour and mosquito transmission has recently been hypothesized.” After the subject — “emergence” — the reader must wade through 25 words before reaching the verb — “has been hypothesized.” Readers will focus too much attention on the anticipated verb to notice the importance of the intervening material.

Swan suggests the following rewrite: “According to a recent hypothesis, virulent Plasmodium falciparum emerged in Africa within the past 6,000 years as a result of a cascade of changes in human behavior and mosquito transmission.” Not only are the subject and verb snuggled together — “Plasmodium falciparum emerged” — but now the important information occupies a key position in the sentence: the end.

The last part of a sentence is what linguists call the stress position. Readers naturally emphasize the information at the end of a unit of discourse, such as a sentence or paragraph, making it the logical spot for new information. Old information does better near the beginning of a sentence, where it grounds the reader in preparation for the mental leap to come. And the more closely the structure matches the reader’s expectations, the more likely the reader is to comprehend what the author is trying to say.

Another mistake often occurs right at the start of the sentence, in the “topic position.” Readers expect to find some sort of bridge between sentences here. If a completely new word or phrase occupies this spot, Swan says, the reader is momentarily confused. For example, a recent paper in Cell begins: “We demonstrate that the tendons associated with the axial skeleton derive from a heretofore unappreciated, fourth compartment of the somites, Scleraxis (Slx), a bHLH transcription factor, marks this somitic tendon progenitor population at its inception, and is continuously expressed through differentiation into the mature tendons.”

The authors unintentionally trip the reader by starting the second sentence with a brand new term — the Scleraxis gene. Although readers can manage this jump, it forces them to divert some of their attention from the science to the writing, particularly if the pattern recurrs. To avoid this, Swan suggests sliding one of the familiar items from later in the sentence to the front to prepare the readers the new information. “This somitic tendon progenitor population is marked at its inception by the gene Scleraxis (Sla), a bHLH transcription factor.” Although passive, the revised sentence smooths the prose, so readers can focus more intently on the scientific content.

Swan’s advice is distributed in a variety of ways. Some institutions, such as JILA, a physics laboratory at the University of Colorado, Boulder, have offices that are ready for editing papers and helping scientists with their writing, and which draw on materials developed by Swan and Gopen. Swan also gives about eight scientific-writing workshops a year in the United States. The Earth sciences division of the Lawrence Berkeley National Laboratory in California has hosted one workshop per year. Divisional director Bo Bodvarson says such training is essential to a successful scientific career, particularly for students whose first language is not English, because clear communication opens doors.

Do the write thing
Despite such enthusiasm, most scientists receive no such training. Part of the reason may be that a significant minority of researchers believe that good writing cannot be taught. Among them is Christopher Miller, a biochemist at Brandeis University in Waltham, Massachusetts, known among editors for submitting clearly written papers and reviews. He says that he doesn’t follow a set of writing rules, but writes instinctively and only when he is in a “writing mood.”

This instinct isn’t something that Miller feels he can convey to his students, so he takes a different approach. Anyone in his lab has two chances to write a paper. “You give me a draft and it will stink, I will write a few things on it and you get another chance,” he says. If that advice doesn’t result in an acceptable paper, Miller writes it himself.

Some students and postdocs aren’t interested, but those who are often produce good second drafts ready for submission, he says. “It has turned out to be productive and fun.”

For those who do not have access to advice such as Miller or Swan, professional manuscript editing services can help. Brian Leonard runs Exact Science Communications, an editing service based in Surprise, Arizona. He says that most of his clients are non-native speakers of English — others want to polish their manuscript to improve the chances of publication. In some cases the suggestion to seek professional help comes from journal editors or referees, he says.

Those editors could also do a lot more to draw better writing from contributors, says Polly Matzinger, an immunologist at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, and a scientific adviser to the Council for the Advancement of Science Writing, which aims to improve the quality of science journalism. Many editors already spend a great deal of time whipping manuscripts into shape, with particular emphasis on the abstract and first paragraph. But Matzinger thinks that journals should push harder and expect good writing in all submissions, possibly even rejecting papers on that basis alone. “Play it up, talk about it, insist on it,” she says. One idea would be for journals to announce an annual award for the best-written paper of the year.

With a lack of big ideas for addressing the jargon problem, bit-part solutions, such as prizes and the range of other ‘Band-Aid’ measures, are currently the best hope for promoting accessibility. Together with the techniques promoted by Swan, they should help to attract scientists to new kinds of abstracts, and keep them hooked until the end of the article. Thanks for sticking with this one. It has a LEX score of 4.1, so you don’t have to be smart enough to read The New York Times to understand it — but it’s probably too complex for cows.

Jonathan Knight writes for Nature from San Francisco.


© 2003 Nature Publishing Group

Pens down Christopher Miller believes that the instinct for good writing cannot be taught.
In the breach.

Now for the subject-matter. You may take a particular

...
If you think that is the case, that open windows are good for health, or that people think open windows are good for health, or that there is a widespread opinion that open windows are good for health, or that it is clear that open windows are good for health, you are probably correct. But you may be misled if you are not aware of some important factors that may affect your perception of the benefits of open windows.

When you open a window, you expose yourself to the outdoors. This can be refreshing and enjoyable, but it can also be dangerous. Exposure to outdoor pollutants, such as particulate matter and ozone, can be harmful to your health. Furthermore, opening a window can让更多 people feel more comfortable and relaxed, which can improve their overall well-being.

If you want to learn more about the benefits of open windows, you may want to consider the following factors:

1. **Air Quality**: Opening a window can bring in fresh air from outside, which can improve indoor air quality. However, it can also bring in pollutants from outside, such as dust and pollen, which can be harmful to your health.

2. **Temperature**: Opening a window can make the indoor temperature more comfortable, but it can also increase the risk of heat loss and energy consumption.

3. **Noise**: Opening a window can reduce indoor noise levels, but it can also increase the risk of external noise pollution.

4. **Privacy**: Opening a window can increase your privacy, but it can also decrease your privacy, depending on the location and time of day.

By considering these factors, you can make a more informed decision about whether to open a window or not. Remember, the benefits of open windows can be significant, but they can also have drawbacks. The key is to balance the benefits with the potential risks.

If you want to learn more about the benefits of open windows, you may want to consider the following factors:

1. **Air Quality**: Opening a window can bring in fresh air from outside, which can improve indoor air quality. However, it can also bring in pollutants from outside, such as dust and pollen, which can be harmful to your health.

2. **Temperature**: Opening a window can make the indoor temperature more comfortable, but it can also increase the risk of heat loss and energy consumption.

3. **Noise**: Opening a window can reduce indoor noise levels, but it can also increase the risk of external noise pollution.

4. **Privacy**: Opening a window can increase your privacy, but it can also decrease your privacy, depending on the location and time of day.

By considering these factors, you can make a more informed decision about whether to open a window or not. Remember, the benefits of open windows can be significant, but they can also have drawbacks. The key is to balance the benefits with the potential risks.

If you want to learn more about the benefits of open windows, you may want to consider the following factors:

1. **Air Quality**: Opening a window can bring in fresh air from outside, which can improve indoor air quality. However, it can also bring in pollutants from outside, such as dust and pollen, which can be harmful to your health.

2. **Temperature**: Opening a window can make the indoor temperature more comfortable, but it can also increase the risk of heat loss and energy consumption.

3. **Noise**: Opening a window can reduce indoor noise levels, but it can also increase the risk of external noise pollution.

4. **Privacy**: Opening a window can increase your privacy, but it can also decrease your privacy, depending on the location and time of day.

By considering these factors, you can make a more informed decision about whether to open a window or not. Remember, the benefits of open windows can be significant, but they can also have drawbacks. The key is to balance the benefits with the potential risks.
A SCIENTIFIC ARTICLE

This page is missing the article text.
Sir Peter Medawar
All these essays are classics of their kind.

Edited by John Maynard Smith

J.B. S. HALDANE

AND OTHER ESSAYS

RIGHT SIZE}

ON BEING THE

A SCIENTIFIC ARTICLE

or even the best possible way. described one way, and I do not claim that is the only way, popular scientific articles in this own way. I have only genuine knowledge while doing so. Everyone must write a certain amount of sell that sort of stuff, and get over a certain amount of so on. I think this is an anti-scientific attitude. But you can with nature. Cheese is a natural food, and better is not. And scientific activities by which we will keep a communication of some concrete actions. Cheese-making is part of the pre-understand project syntheses, not the extreme specificity of show cheese as part of the Mycological Universe. We do not the jokes about the smell of cheese.

The jokes from Epides' Cypers, if for a lowprow, any of

160
Style, hyphens and the split infinitive

Authors of scientific papers always await referees’ reports with trepidation. Nothing can be as disheartening to authors as a summary rejection of their manuscripts by editors, who often base their decisions on devastatingly worded comments from ‘peer reviewers’. Many referees seem to delight in dismissive language; anonymity provides an impenetrable shield. Sometime ago, I received referees’ reports on a paper that I had submitted to a journal published from the United States. The editor’s covering letter seemed promising, requesting the inevitable revision. But one of the reviewers, while grudgingly accepting that the mundane science described in the paper was publishable, went on to launch a vigorous attack on my language. He (and here I speculate on the sex of the referee) charged me with ‘violating’ two of his ‘pet peeves’. First, he came down heavily on my use of the ‘split infinitive’. Second, he was pained by my use of ‘nouns as adjectives to modify nouns that are used as adjectives without hyphens’. He did add, somewhat condescendingly that it was all a matter of taste, implying clearly that while we might even agree on matters of science, our differences on points of grammar were irreconcilable. Upon re-reading my own manuscript I realized that I could neither recognize split infinitives, which were undoubtedly strewn all over the paper, nor could I decide on the location of the missing hyphens. Stung by the criticism I refrained from submitting a revised version of the paper, waiting for an opportunity to clear my increasing doubts about my own grammar. At one of the symposia where booksellers display their wares, I came across a copy of *Scientific English* by Robert Day (Universities Press, 2000). Reading a book on writing scientific papers is not easy. But, Day’s preface was both encouraging and entertaining. He quoted both Plato and Confucius. The former had said: ‘Beauty of style and harmony and grace and good rhythm depend on simplicity’. The latter was characteristically brief: ‘In language clarity is everything’. A little further on Day came to my rescue rather forcefully: ‘I have good news. You may split infinitives. In fact, you may, on occasion, violate every one of the “rules” dreamed up by generations of grammatical fussbudgets. … The obvious purpose of grammatical rules is to facilitate clear communication. When rules of grammar do not serve this purpose, they should be disregarded’ (p. xi). On the hyphen, Day was equally encouraging: ‘The hyphen has a number of uses, most of them confusing’ (p. 92). For good measure he concludes by quoting John Benbow: ‘If you take the hyphen seriously, you will surely go mad’. Even *Fowler’s Modern English Usage* (ELBS and Oxford University Press) is harsh on the hyphen, quoting Winston Churchill’s famous dictum: ‘One must regard the hyphen as a blemish to be avoided as far as possible’. In protesting against the hyphen, Churchill argued ‘that you may run them together or leave them apart, except when nature revolts’. Indian scientists (and undoubtedly scientists from many non-English speaking parts of the world) are often admonished by British and American referees: ‘The authors should have their manuscripts read by a native English speaker’. This kind of comment is received even when the language is passable, often motivated by a casual reading of the authors’ address, rather than by an analysis of the manuscript’s grammar. Despite my current aversion for grammar, stimulated by a disagreeable referee, I must confess that as a reader and editor I often wish that manyauthors paid some attention to writing style. There is nothing more disheartening for an editor than to receive a letter for publication, which is completely incomprehensible. Summary rejection of such contributions seems unfair; improving them by editing seems impossible. This journal’s office appears to be receiving an increasing number of manuscripts, which are very poorly written, carelessly proof-read and often display a complete disregard for the suggested format. After years of glancing through manuscripts in diverse fields of science, I have reached the inescapable conclusion that compulsory instruction on writing scientific papers might be an invaluable exercise for research students. The habit of consulting dictionaries and style guides is to be encouraged and analysis of the text of published papers, for the manner of their construction, might be helpful. Interestingly, even as I was thinking about grammar and style a book entitled *Communicating in Style* by Yateendra Joshi was received in our office for review. This reasonably sized (about 250 pages) and moderately priced (Rs 300) book, published by The Energy and
Resources Institute (New Delhi, 2003), appears to have been generated for in-house use. But, this book is not for those interested in clarifying the nuances of grammar; it seems more directed towards those interested in presentation. Curiously, the cover carries a blurbs attributed to John le Carré: ‘A gem. Courteous, unfrightening and essential. A perfect companion to Fowler’s A Dictionary of Modern English Usage for today’s communicators’. I would prefer Fowler and common sense. A book more directly addressed to scientists is The ACS Style Guide (Dodd, J. S., ed., American Chemical Society, 1997). Although this book focuses primarily on chemistry, it would undoubtedly be useful to writers in all branches of science. Most satisfying, I found both the split infinitive and the hyphen discussed. The ACS committee decreed: ‘It is acceptable to use split infinitives to avoid awkwardness or ambiguity’. Hyphens are elaborately treated, as the literature of chemistry is replete with broken words; scanning through long lists of ‘ unit modifiers’ (two words that together describe a noun, e.g. electron-diffraction, excited-state) I could not help but feel that the ACS had ignored Churchill’s dictum. The ACS Style Guide is the kind of book that ought to be lying around in our laboratories, in the hope that students may idly turn its pages and in the process come across useful nuggets for improving the style and presentation of papers. There are sections in this book which go beyond writing style, providing useful tips on copyright matters and oral presentations of science. In these days of facile internet access and rapid ‘downloads’, many seminars are delivered in which figures, data and even animated movies of real and imagined interactions of atoms, molecules, cells and organisms are projected. Most often, there are no references to the site from which the illustrations have been taken. Computers, the ‘world wide web’ and ‘Powerpoint’ have made plagiarism of sorts a routine event in many presentations, both oral and written.

While the paper written for a professional scientific journal can get away with a significant degree of obscure language, articles addressed to a more general audience must be clearer. This journal, for instance, has a section entitled ‘general articles’, which should hopefully be accessible, at least in some part, to a non-specialist reader. Unfortunately, most submissions in this category are manuscripts which are unintelligible; their sole claim to being ‘general’ is the absence of original research normally found in a scientific paper and the lack of scholarship that one associates with a ‘review’. In an essay on How to write a popular scientific article, J. B. S. Haldane provides compelling advice: ‘The first thing to remember is that your task is not easy, and will be impossible if you despise technique. For literature it has its technique, like science, and unless you set yourself a fairly high standard you will get nowhere… You must … know a very great deal more about your subject than you put on paper. Out of these you must choose the items which will make a coherent story’. For articles that do not have to be published immediately, Haldane has an interesting prescription: ‘When you have done your article give it to a friend, if possible a fairly ignorant one. Or put it away for six months and see if you still understand it yourself. You will probably find that some of the sentences which seemed simple when you wrote them now appear very involved. Here are some hints on combing them out. Can you get in a full stop instead of a comma or a semicolon? If so, get it in. It gives your reader a chance to draw his breath. Can you use an active verb instead of a passive verb or verbal noun? If so, use it’ (On Being the Right Size, J. B. S. Haldane (ed. Maynard Smith, J.), Oxford University Press, London, 1985).

The growing technical complexity of the scientific literature makes even ‘reports of research focused on an outstanding finding whose importance means that it will be of interest to scientists in other fields’, impossible to comprehend at first glance. I took the above quote from Nature’s guide to authors (cf 2004, 427, 84). The provocation was my attempt to read a letter entitled ‘Bcl10 activates the NF-kB pathway through ubiquitination of NEMO’ (Zhou et al., Nature, 2004, 427, 167). Undoubtedly important, the paper illustrates the difficulty that non-specialist readers face in reading papers in many disciplines of science. Abbreviations and acronyms abound. Some have been introduced into the literature to satisfy a momentary humorous impulse of an author. NEMO took me back to Jules Verne and Twenty Thousand Leagues Under the Sea. The literature of molecular biology is full of genes and mutants with curious names—hedgehog, sevenless (and, of course, son of sevenless) and even an apparently familiar, laloo. Biology’s literature is rapidly becoming as intractable as chemistry, with nomenclature providing an impassable obstacle for a general reader.

Writing style is, of course, a ‘matter of taste’ as my referee correctly pointed out. Grammar can serve as a guide, not a shackling constraint. Ultimately, a published paper must be useful. For this, it must be understandable. A little attention to style and technical crafting while writing a letter, a paper or a report would help readers (and copy editors). Reminding authors of the rules of grammar, on occasion, may be valuable. I, for one, will look with respect at both the hyphen and the split infinitive, in future.

P. Balaram