A few semantic and technical paper writing tips

One of a series of time-dependent evolving 'advice' essays that I will intermittently post and intermittently update on my web site.

Introduction

Good writing is important. No one is perfect; I don't pretend to be, but I try hard. I have been drawn increasingly to this issue since moving to the USA and expending much effort editing the writings of my students and postdocs and reading, editing and refereeing far more papers and grants. I admit to finding it distracting—sometimes very much so—if I have to re-read sentences to resolve ambiguity in their meaning. Punctuation and style are important. If you want people to read your papers, make them readable. My perception is that writing has declined in quality over the years whereas mine was not so good long ago but, as I have become more aware of it, I try to improve (but please do not expect these web commentaries to have received as much effort as the printed word might have).

There are many guides to good writing, grammar, syntax, etc. Not all of them agree on some points. I will note a few important and frequently misunderstood issues here, which will sometimes mean I have come down on one particular side about something that all authorities do not themselves agree upon. There are online sources as well as books. For the latter, my favorite remains 'The *New* Well-Tempered Sentence: A Punctuation Handbook for the Innocent, the Eager, and the Doomed' by Karen Elizabeth Gordon (1993), which has been around far longer than Lynne Truss's (2003) best-selling 'Eats, Shoots & Leaves: The Zero Tolerance Approach to Punctuation'. My other recommendation, from among the large number of offerings, is The Economist Style Guide: 9th Edition (2005). All three have the advantage of being amusing to read as well as informative. For easy access, there is an excellent and comprehensive online Guide to Grammar and Style by Jack Lynch, Associate Professor in the English department at the Newark campus of Rutgers University http://andromeda.rutgers.edu/~jlynch/Writing/index.html.

One of my rules of thumb is that the more people who are acknowledged for 'critical reading of the manuscript' the less likely it is to be an easy read. Writing is the responsibility of the authors. Usually there are several, so one hopes that at least one has been taught how to write English. Non-native English speakers should ask one who is native and a good writer (it is not an automatic connection) to spend some time helping them, probably for more than just a bland acknowledgement—or perhaps even a beer (or how about helping them with an experiment, for example).

Specific 'nags'

Data, bacteria and sera are plural: the singular are datum, bacterium and serum. The OED actually gives serums as an acceptable plural form! I may be fighting a losing battle with data—but why not abolish singular forms of other words. After all, it appears that politicians, industrialists and writers is each capable of degrading other niceties that is in our culture. Why don't we all talk like Ali G!

Commas and semi-colons are the most useful—but most often omitted and abused—tools of punctuation. The easiest rules are to imagine where pauses would be inserted if reading a sentence, and to remember that phrases containing subsidiary information, meaning that it encompasses clarification, etc, would normally be flanked by commas.

Sentences should be clear and not too long. In contrast, sentences can sometimes be advantageously combined.

Do not use 'since' when you mean 'because'. Since implies temporal distance, as in "It is difficult to remember small things since I was diagnosed with dementia, because that is the nature of the disease".

My current nominee for the most over-used unnecessary word is 'different'. It might be appropriate to say "we got different results from different experiments" so long as they were different *types of* experiments. It is not appropriate to say " we performed five different experiments" when really you performed five replicates of the same experiment.

One of the most over-used unnecessary phrases is 'in order to', as when we say "We performed such and such an experiment ~~in order to~~ investigate the effects of long sentences on readers' attention span".

Avoid frequently starting sentences with words like surprisingly, strikingly, unexpectedly, as expected, etc. It is not good style and it can become quite annoying when overdone. It is usually better to state the result and then make a statement of why the results were surprising, striking, etc, together with the reasons.

Compounded words used adjectivally (what is the correct term for this?).

For example, I will base my theory on wild-type cells but I will report my observations on the wild type. I will use probes that are 100 bp but, if I have to, I must refer to them as 100-bp probes.

Things can be present in but only absent from.

There are many words that cannot be qualified: your cells cannot be completely dead, partially eliminated or partially identical. Your genes or proteins can be 50% similar but not 50% homologous.

And, a few comments from George Carlin ('Jammin in New York' show, December 2007: hilarious)... "people add words when they want to make things sound more important". A few of his examples from airport routines... "start the boarding process (start boarding); responding to an emergency situation; complete stop; final destination (all destinations are final); and — perhaps more relevant to us — pre-heated, pre-recorded, pre-existing. I am sure we can all think of others.

That and which

Why is it so difficult (or is it just that, like me, you were never formally instructed in such things)? Although some contemporary style pundits say it doesn't usually matter, scientific writing could be said to be THE place where IT DOES matter.

One paper I was reading recently managed to use 'that' and 'which' in the opposite of the rule (for example; "...was used to transfect *T. brucei* 29-13 cells that express the tetracycline repressor as well as the T7 polymerase...". BUT, all 29-13 cells have this property, so this is not a defining clause. A second example; "...during silencing, epitopes, which are normally internal, were exposed, like in apoptotic cells...". This construction implies that all epitopes are normally internal, whereas the author meant the specific class of epitopes that are normally internal. My apologies to the authors, if they recognize themselves, but I just happened to be reading your paper at the time I started writing these notes — an unhappy coincidence! The discussion section of this paper also provides good examples of the lack of or misuse of commas.

Another example: "Despite its [the specific leishmania enzyme the authors were referring too] shared evolutionary ancestry with the short chain dehydrogenase family which includes DHPR...". In this example, because I do not know anything about the supposed "short chain [short-chain] dehydrogenase family, specifically whether it is one family or several, I cannot determine whether the subject is a member of the single family, in which case 'which' should be preceded by a comma, or if it is a member of one particular family of short-chain dehydrogenases, in which case it should be the family that includes DHPR.

I noticed another good example of incorrect use in the NYT obituary (January, 2000) for Paul Sigler, where it said that "... he determined the structure of the RNA molecule, which initiates protein synthesis."

It should have said that "... he determined the structure of the [specific] RNA molecule [tRNA] that initiates protein synthesis.", because he determined the structure of a specific RNA molecule, there being many, not the one that 'which' implied!

Technical tips

These days, papers are exclusively typeset (if they will in fact be printed on paper at all) from the corresponding author's electronic file. This has placed the onus on authors to properly format their paper from the outset (they are probably typing it themselves too, these days). Doing this properly can be tedious at first, but has benefits later. Firstly, it makes proof-reading less onerous once your paper is accepted; secondly, referees can be distracted by badly formatted papers (they will increasingly be reading them on

screen, in pdf format); thirdly, most journals allow you to post or circulate preprints of the original version of your paper and these will look nicer if properly formatted.

So, check a paper in a current issue of the journal you are submitting to determine the styles (normal, headings, fonts, etc) that you should use. It looks much more professional if the original submission is properly formatted, and it makes it easier for referees to read.

Most of us use only 10% of the facilities of our computers and applications. This is the time to save time by learning a few basic technical aids.

Most people will be using Microsoft Word and EndNote (it's a pity the Mac-centric Niles Corporation, who invented EndNote for the Mac, sold out to the apparently PC-centric Thompson megacorporation) or Bookends (my personal favorite — more powerful, much less pricy, and more customer-responsive — http://www.sonnysoftware.com) to write their papers. With successive upgrades, word-processing programs become bloated with new features, most of which you will never use. I am just going to suggest a few things that will save you huge amounts of time. The time to learn these few points will probably be saved in the first paper you write.

All my suggestions are based on the Mac OS 10.5 versions of these two programs—but Mac and PC manifestations of Word are similar and produce largely interchangeable documents. I currently use Word 2008 (with all its updates and patches) and Bookends.

Word: the must-do things

Line spacing and justification

Most journals required submissions of manuscripts that are double-spaced, left aligned (most ask NOT to justify) and pages should be numbered (reviewers hate un-numbered pages: numbering lines can also helpful to referees, but few journals insist on this).

Use Styles

This means that you will assign a 'normal' style for the body of your text and assign styles to headings. There are huge advantages to using styles rather than giving ad hoc font formats to different parts of your paper; first, if you use special fonts for symbols, etc, they are preserved if you re-define the format of the normal style (which is not true if you select text to reformat the font); second, if you have to reformat your paper for a different journal or to make a preprint, it is easier to redefine the styles rather than go through each paragraph or heading and correct it. Add spacing before and after headings, in their style definitions, so you don't have to put extra lines between headings and text. Use double spacing as part of your 'normal' style definition.

You can define a default document as Word 'Stationery', so you don't have to redefine styles each time you start writing a new paper or document.

Use the 'track changes' tool

This is immensely helpful when different authors will revise the manuscript. You can turn the 'highlight changes on the screen' and 'highlight changes in the printed document' off, if there are many changes and the screen starts to look confusing while you are writing. Your co-authors can turn this option back on to quickly see what's changed. To avoid mounting confusion (and the software itself can become confused) it is usually best to review a previous author's changes and 'accept' them, before adding your own. Tracked changes can be hidden but comments left visible, which can help distinguish editorial changes from messages or comments between authors.

Spelling and autocorrect tools

Use the real-time spelling checker AND make a custom dictionary to which you easily add words that are not in the standard dictionary. Keep this dictionary where it will be backed up. It is well worth the small amount of time learning how to do this, and setting up the 'autocorrect' tool as a shortcut for typing terms that you use repeatedly. For example, you can define a shortcut that will be automatically 'corrected' when you type it. For example, if I type trbx, *Trypanosoma brucei* is inserted formatted in italics. One can create a large number of such shortcuts—even entire sentences or blocks of text, like your address, for example. A good idea is to make a document in which you first type every shortcut (organism names, for example) that you routinely use, format them, then assign an autocorrect 'abbreviation'. There are other uses for this tool, too.

Know your keyboard commands

There are many characters that can be invoked from your keyboard, besides the obvious ones. Some of these are universal but some may be OS-specific. A must-have extra on my Mac is a small shareware application called PopCharPro. This is a pop-up panel that shows all the characters available in the current font or any font on the system. Alternatively, use the character palette on the Mac menu bar.

There are some must-use non-keyboard characters. For example:

µ is option-m on a Mac.

Use the degree symbol ° for temperature (shift-option-8 on a Mac).

Use the prime (´) sign where appropriate (shift-option-e on a Mac).

When giving a range of values, the correct symbol to use is NOT a hyphen but an n-dash (–), which is option-hyphen on a Mac. An M-dash is shift-option-hyphen (—) and is used to flank a phrase — like an aside — that might otherwise end up in parentheses.

With PopChar Pro you don't have to remember this: clicking on a character in the popup panel will automatically insert it into your text. The only problem with some add-on programs like PopChar or Default Folder is that they often work with all programs except Word, because Microsoft insists on doing things that contravene system programming guidelines.

Bibliography tools: EndNote and Bookends

Set up shortcut keys for reference insertion, etc.

Ensure that the reference is correctly formatted in EndNote or Bookends . This will save you lots of time when re-using the reference or reformatting a manuscript! Especially make sure titles are correctly formatted, including the use of italics for species names. Learn how to specify the correct journal abbreviation for compiling your bibliographies.

In my experience, most of the Journal Styles that come with EndNote need tweaking to make them exactly right. Check the reference format in a current paper in the journal you are submitting to: journals sometimes change their reference formats. Bear in mind that there are multiple options you need to select to get perfect citation and bibliography formatting.

When you cite multiple references to reference a single point in the text, make sure EndNote or Boookends style is set to cite them (in the text) in date order, not alphabetical. I think this is the appropriate/courteous way to list discoveries, etc.

DO CHECK that the bibliographic program has captured all citations that you placed in the text. The programs are perfect, but sometimes you will have inserted a duplicate reference into the library, or deleted the field designation, or inadvertently changed the reference number that EndNote uses to track citations.

Figures: some very basic things

Check the journal's requirements for the preparation of print figures before preparing them! Most journals do not accept PowerPoint files for print figures. Most have very definite specifications for print figures, some examples of which are—

Print figures normally require the use of CMYK colors, not RGB. Online-only journals will use RGB.

If you do not have color, and to avoid being charged for it when you didn't expect to be, set the Photoshop workspace to GREYSCALE. The files will be much smaller.

Resolution usually needs to be 600 dpi for 'photos', color or greyscale and 1200 dpi for line drawings, but check individual journals.

Some journals request the figures be submitted in the native file format: others specify more generic formats like TIFF or EPS. These formats have very different properties and are not randomly interchangeable. Illustrator files that consist of vector drawings and text should be saved as EPS and Photoshop as TIFF. However, Illustrator figures that consist mainly of imported TIF files should be saved as TIF (but why are these in Illustrator rather than Photoshop). Greatly reduced file size can be obtained by saving Illustrator or Photoshop files that consist mainly of photos plus legends as JPEG. This is probably the simplest solution when a journal (like NAR) asks for all figures to be included in one manuscript file (with the text) for initial submission.

Photos, based on pixel graphics, should be prepared in Photoshop. Legends can easily be added in Photoshop, in the text layer. But, do not flatten/merge the layers until everything is perfect at the final required image size because, once merged, text looses its special properties.

If you use Illustrator, and you need to import image files, these MUST be cropped in Photoshop before importing. This will save a lot of time, hassle and space on the Illustrator workspace. Images can be resized in Illustrator, but not cropped.

Do not leave lots of blank space between different sections of figures.

Do not clutter figures with unnecessary objects or text. Make sure text is all in the correct font (normally Helvetica, not one of its varieties) and suitable size. Use a size that is slightly larger for the top-level labels, usually A, B, C, etc. Do not put brackets or periods around or after the letter.

Always test what the final figures will look like by printing them at the likely actual size, which will be 1 or 2-column widths of the journal.

Do not put figure numbers or any extraneous text that should not appear in the printed version in the original files!

DO put a figure number on the pdf versions, for the reviewers, when making a composite pdf file, or when the journal web site makes it for you. Numbers must not be included in the production version.

Submission

Turn on Tools>Track Changes>Highlight Changes>Highlight Changes on Screen, to check if changes have been tracked, in which case they will all be apparent in the raw document (and possibly even in a pdf version) if anyone else (a referee, for example) might get access to the raw word file, which could be embarrassing!

If there are changes, go to Tools>Track Changes>Accept or Reject Changes, and select the 'Accept All' option, which will do as it says and delete the track chances information. Save this version with a new name to avoid over-writing the original.

Select the entire document and use the Tools>EndNote>Remove Field Codes option to remove EndNote field codes (or use Word Unlink Fields command). DO THIS ON A COPY OF THE DOCUMENT THAT WILL BE USED FOR SUBMISSION. KEEP THE ORIGINAL FILE with the field codes, in case you have to recompile the references later.

If at all possible, submit your manuscript and figures for review purposes as a single pdf file. Then you have complete control over the appearance and scaling of figures, etc. Most journals, however, seem to prefer that you submit raw files that they convert to pdf, which can sometimes cause problems.

If you saved your figures as pdf files, try to ensure that the page setup is portrait, so a referee will not have to turn the pages sideways to read them!