

Overview

Early on in graduate school, my collaborator advised me to read the full Strunk and White book on Grammar. I am not a perfectionist, but this experience has stuck with me, and I believe that it is important first to learn the rules of grammar and then to learn when they can be broken.

This is a summary of the rules for scientific writing that I see broken most often to the detriment of the paper.

Unhelpful words and phrases

The following words make it hard for me to read and understand sentences.

- etc.
- ... (except in mathematical contexts)
- and/or, or any use of the backslash
- &
- vs. (write *versus* instead)
- “utilize” (write *use* instead) – use the shortest, most common word that says what you mean
- contractions with apostrophes (e.g., “don’t”)
- “eg.” (write “e.g.”, instead)
- commas go before “which,” but not “that,” as modifiers of nouns
- think carefully when adding or omitting a comma. Too many or too few hurt the sentence flow.
- avoid all adverbs: “very”, “really” especially
- negative + verb: e.g., “not enriched” should be “depleted.”
- always use an Oxford comma: “A, B, and C” not “A, B and C”

Sentence structure

- Don’t write long sentences. Break up long ideas with periods, semicolons, or triple dashed phrases (e.g., “..., with which gene A—when measured in the unexposed samples—does not appear to interact”)
- Prefer active and not passive voice
- Use simple words, and simple sentence structures.

Semantics

I do not know if it is possible to teach a person to write meaningful, clear sentences and paragraphs. I think the phrase that another collaborator said once that I repeat to myself often when writing papers is:

Words mean things.

It sounds simple, but it embodies the idea that you should not use a word that does not capture the object, idea, or action you are trying to convey to the reader. For example, “We collected GWAS data for 943 participants.” What does it mean to collect “genome-wide association study data“? Do you mean that you collected genotype data for these participants? If so, say that.

Logical flow is another hard concept to teach. Paragraphs read almost like proofs. Avoid writing things until all the required information leading up to that statement have been written.

Finally, order of presentation of ideas is also a difficult and subjective area. I look at the section and subsection headers to see if there is logical flow to the whole document. These section and subsection names are very helpful to directing the reader as to what is coming and what is in that section, like code comments. These should also be as descriptive as possible without being unreadably long. Some writers advise to write complete sentences as the section names with the punchline: “Differential expression results suggest widespread impact of exposure to statins.”

Detailed outline

I advise students to put together a paragraph-resolution outline of their paper before actually writing, including figures and potential caption first lines.

How should you do this? Look through the related work in the journal to which you are planning on submitting your manuscript. Often, there is a published paper or two that are very well written and organized. **Without using any of the words or phrases that the other paper does** (this is plagiarism) follow their organizational lead in terms of presentation of ideas and details presented; adapt to the context of your research. If you are writing a Methods paper, do not use the outline of a Research paper, and vice versa – paper type matters.

Really finishing a document

You have sent your advisor the draft, received feedback, and are ready to submit! At this point, there is (I find) as many as 10 hours left of work before actually being able to submit a journal length paper; conference papers are somewhat

easier. Here is my general checklist (parts adapted from Dave Blei); some of this advice assumes you are using TeX.

- Authors names and affiliations (out for double blind review, double-checked for journal)
- Spell check!
- Each bibliography entry should be two lines. (there are exceptions.)
- Make the bibliography consistent, e.g., first initial and last name only. omit “Proceedings of Conference” and just say “Conference”, e.g., “Artificial Intelligence and Statistics”. Omit page numbers if they send you to a third line. Capitalize proper names in curly brackets like Bayes and Dirichlet.
- Review every single bibliography reference for proper information, not missing authors, etc.
- Ensure that the colors, ordering, and inclusion of methods, experiments, and results are consistent across Methods, Figures, Tables, and Results.
- Read over Methods again. Make sure that you could repeat each step without having to go to your code. Also make sure that this is boring recipe book writing, and not motivation or explanation.
- Look for paragraphs that end in 1-3 orphan words and edit them to save a line. This makes the writing better and saves space. And its a nice (and random) way to edit your paper.
- Captions should be clear and consistent; they should begin (or end) with a sentence saying what you want the reader to get out of them (more information in the FIGURES document).
- use Ref. [x] for when you need to use citations as a noun. alternatively use the author names, but this might take more space. But try to avoid using citations as a noun.
- Instead of “Equation“, use”Eq.“; Instead of “Figure” use “Fig.“; instead of “Table”, use “Tab.”
- Write a draft of the cover letter (for journal articles)
- Read over the formatting instructions for the journal or conference, including figures. Do exactly what they say. Except for: put figures inline, since, as a reviewer, this is always better.
- Make sure each labeled equation is referenced in the document. Make sure each figure and table are referenced in the document at least once. Make sure the order they are referenced are the same as their numbering. Do the same for the supplemental data.

- Make sure the supplemental data, code, etc. are all ready to go. Have code ready to distribute to reviewers, even if it is a tarball of the code that you can upload to the journal to distribute to reviewers.
- Look up a set of people as reviewer suggestions for the paper. Clear this with your co-authors (and ask for their suggestions). Also think about potential reviewers with conflict of interests to exclude (I generally only exclude current advisors or collaborators, if they are obvious reviewer choices).
- Read, reread, read again, read out loud
- Spell check again!
- Get a (suitably journal-anonymized) version ready for ArXiv or BiorXiv, after discussing with your advisor.