Things I Keep Repeating About Writing – Claire Le Goues

10-12 minutes

I often write papers with students, or read students' papers to provide comments, and I find myself saying the same things over and over, especially the first time out.* So: here's a blog post I can point them to to (hopefully!) save us all some time and trouble. I plan to update it as I remember more things I say repeatedly.

I'm happy to argue these points, and take suggestions to expand the list. I'm not claiming that I'm the world's foremost writing expert, and some/many of these are the product of relatively arbitrary preference. But, (A) this is targeted first and foremost at my own students, so my preferences matter, and (B) I'll try to justify when I can.

This list isn't a complete delineation of all the rules of English grammar. Follow those rules too, even if they're not on this list.

* These are not the kinds of comments I typically make when reviewing, where I focus less on style.

Use clear and precise language.

Use short, declarative, active sentences. BANISH THE PASSIVE VOICE. If you went to an American high school you probably need to retrain your instincts.

Use adverbs and pronouns judiciously:

- Adverbs are often imprecise: what does "incredibly" add to the phrase "incredibly important" that the word "important" lacked on its own? How much more important than important is something that is incredibly important?

- Pronouns are often unclear with respect to their antecedents, which can confuse the reader.

Be as explicit/concrete in your statements as you can. This is perhaps best illustrated by example (courtesy Yuriy Brun): Instead of "The dataset has a few attributes.", say "The dataset has 22 attributes." Avoid descriptors like "a number of" or "several", which rarely add meaning. Instead of "We

performed a number of experiments." or "The cat had a number of lives.", try "We performed four experiments.", "The cat had nine lives."

(To highlight the point, consider the sentence(s) without "a number of": "We performed experiments."/"The cat had lives." See how the meaning didn't really change?)

Related: do not use more syllables than necessary.

Two easy manifestations of this rule are the following transformations that can be applied universally to your draft:

- "In order to" –> "To"

- "Utilize" –> "Use" (unless in the context a discussion of CPU utilization, where it's reasonable).

The point of writing is to communicate an idea. Using more syllables than necessary obscures the idea without adding meaning.

Present numbers properly.

Write out in letters all positive numbers less than or equal to 10, unless they are in a sentence with a number greater than 10 (ETA: like 110, which makes this sentence comply with my rule). I don't know why.

Right justify columns of numbers. I will repeat this in all-caps, because I really mean it: RIGHT JUSTIFY COLUMNS OF NUMBERS. Ensure that the correct number of significant digits are used (your stats package is giving you waaaaay more than is appropriate), and that decimal points align.

You will argue with me about this, because you really want to left-justify or center them. I don't know why. A reader should be able to quickly scan a column of numbers to get a sense of magnitude, and cannot do that if they are left-justified unless they are all (coincidentally) the same order of magnitude.

Text in columns should be left justified. Never center anything that's not a column header.

Typesetting/copy-editing minutiae.

(On all of these, the answer to Why? is usually: Because.)

Capitalize Table, Figure, and Section. Refer to sections only, never subsections, even when you're referencing an actual subsection (e.g., Section 4.1, not Subsection 4.1). Include a non breaking space (~) between the words Figure/Section/etc and the \ref.

Capitalize and punctuate section/paragraph headings/captions consistently. If one ends with a period, they all should.

Do not use citations as nouns. No: "In [14], Hazelwood et al. describe facts."

Yes: "Hazelwood et al. [14] describe facts." (H/T Kim Hazelwood)

Citations go before punctuation, with a non-breaking space between the word and the citation. Footnotes go after the punctuation, with no space.

An em-dash is three dashes in latex. You use these to offset text, like a parenthetical but without parentheses (I'd give an example but wordpress converts my triple-dash into an em-dash automatically so it's hard to see!). An en-dash is two dashes and is only used for ranges (like page numbers). A single dash is used in hyphenated words. You probably don't need to hyphenate compound words nearly as often as you think you do. No spaces around dashes; sometimes a space after a hyphen, depends on the circumstance (e.g., pre- vs. post-condition).

(The actual rules for dashes and hyphens and compound phrases are complex, so beyond that I'll punt to another website instead of typing them all out.)

Abbreviations should include appropriately placed periods, that is, after every shortened version of a word. So "also known as" is abbreviated "a.k.a."; versus is abbreviated v. or vs.; "et cetera" is abbreviated "etc." (a mistake I made in the first draft of this document!). Et al. is another one, and a pet peeve (period after the al., which is short for *alii*, not the *et*, which just means "and" and isn't shortened). Et al. should not be italicized, though I took some convincing on this. It should be separated from the preceding name with a non-breaking space.

Always put a comma after i.e. and e.g., and use them properly (i.e. means "put differently" or "in other words", e.g. means "for example").

It didn't initially occur to me to include "use the Oxford/serial comma," because doing so is so obviously correct.

Make your figures and tables maximally readable.

Do not hit the page limit by shrinking your tables and figures. Assume your reader is old, blind, lazy, and also colorblind. Print out your paper at least once on physical paper and make sure you can read the figures and tables. I do actually complain about this when reviewing.

Choose colors for graphs and figures that show up when your paper is printed in greyscale. Go to http://colorbrewer2.org/ and choose "colorblind safe" and "print friendly" to find color combinations that work.

Use booktabs for tables. They look so much nicer and internal rules do not actually increase readability.

The default font size for labels on graphs coming out of basically any package (Excel, R, etc.) is too small. Don't let the defaults boss you around.

Use latex, bibtex, and version control in a way that makes your advisor

happy.

There are myriad differing opinions on this; of all the "rules" on this page, these are almost certainly the most CLG-specific.

Naming. <mark>Name your .tex file (and project/directory) something more informative than "paper". Reasonable schemes include but are not limited to: "lastname-projectname-year",</mark> "projectname-venue-year", "lastname-venue-year".

Version control. I prefer to collaborate using git, mercurial, or svn, through a hosted repository. github or bitbucket are fine. My username basically everywhere is clegoues.

Do not check in byproducts of the build process, *including the PDF.* If you do, we will conflict every time we commit, which is annoying.

Because I like to use git/hg/svn, I strongly prefer hard line breaks throughout a document. My editor default is 80 characters. Fewer than that is fine; longer gets silly. Some people like to line break at the end of sentences, which I think is weird but preferable to no line breaks at all. Note that I don't "rewrap" unless things get crazy. The point is just that if lines are roughly 80 characters, line-based diffing and merging (as done by git/hg/svn) works pretty well and simplifies collaborative editing. If paragraphs are all one long line, merging becomes substantially more difficult.

Tools. I prefer to write papers in emacs and will add a Makefile to your directory, and then build the paper using "make". You can do the same, or use whatever other editor/tool you like.

I tend to dislike shared latex editing sites like sharelatex, but make allowances, especially when there are fewer than three collaborators. I prefer those options to emailing a Word document around. I prefer *that* to those WYSIWYGs that generate latex, which I won't use. Google Drive is OK for early drafts, but I'd generally prefer we just skip to the latex.

Latex. I prefer latex documents to be structured as "all one file" rather than having sections or subsections in multiple latex documents and inserted via \input. Dissertations/theses are a reasonable exception. I compromise on this based on the preferences of my colleagues, but given a choice…

Leave space/a subsection/a paragraph for acknowledgements at the end so we can acknowledge sponsors without having to panic to make space right before the camera ready.

Bibtex. Give your bibtex entries reasonably indicative names. If you cut and paste it from the web somewhere, ensure that it's done properly (some sites make everything a @misc, which is almost always wrong) and modify the bibtex so that it's reasonable. Definition of reasonable: special characters

are copied properly; authors names and title are spelled/capitalized correctly (don't forget non-breaking spaces where relevant, like in your advisor's last name…); includes venue, preferably spelled out along with its acronym, but you can drop the "Proceedings of the 23rd Annual ACM/IEEE blah blah" in favor of just "International Conference on Software Engineering"; includes year and page numbers. The rest is mostly optional.

Fair warning: I tend to insert broken bibtex cites as I write to remind myself/you to put references in appropriately.

(Shout out to the numerous others who commented/made suggestions/nit-picked my own copy editing, with especial thanks to Kim Hazelwood and Yuriy Brun*, two of the only computer scientists I've ever met who are bigger sticklers than I am on grammar/typesetting.)*