Tips and tricks when writing papers

16 October 2012

Andrew Hopkins

Introduction

This document is a collection of tips, advice and suggestions for people preparing astronomical papers, including many common latex mistakes and corrections. This list has been compiled following revisions of many papers written by students and colleagues, and having been exposed to a large number of errors that appear to be most common.

Common LaTeX errors

1. Use the `\label` command to mark Sections, Figures, Tables, etc., and the `\ref` command to refer in your document to Section, Figure, Table, etc., numbers. This should ensure that you will never have incorrect references anywhere. Also, use a bibliography and the `\cite` command (and aastex and other variants) to refer to references. This ensures that all required references appear in the bibliography, and all bibliography references are cited in the text.

2. An apostrophe is not an inverted comma. Use the forward-quote symbol ` to get an open-quote symbol: '. Using the standard quote-symbol, the apostrophe, ' gives a close quote *only*: ' . If you do this, your 'phrases' are not properly 'enclosed'. Also, inverted commas should be used, two forward-quote symbols matched by a double-inverted-comma ``xxx" by preference: "Why is a raven like a writing desk?". (Two apostrophes to close the quote works the same way as a double-inverted-comma.)

3. Spacing. Some special characters, commonly the Angstrom symbol `\AA`, do not have an implicit space after them. This leads to no white-space being inserted after the symbol as in: My ruler is 3×10^9 Ålong. Source here is: `My ruler is $3\times10^9\,$\AA long.` This can be rectified in two ways, including the symbol within the math-mode dollar signs, or by telling latex to leave a space after the Å symbol with a closing backslash: `\AA\`. The source `My ruler is $3\times10^9\,$\AA\ long` gives: My ruler is 3×10^9 Å long.

4. Spacing. Use a small-space, obtained with LaTeX's `\,` character, to separate units from their quantities. Compare the following. `$10\,$m` gives 10 m. `$10\ $m` gives 10 m. `$10~$m` gives 10 m. Other poorer constructions, not to be used, include: `$10 m$` gives $10m$. `$10 $m` gives 10m. `$10$m` gives 10m.

5. Spacing. A backslash followed by a space produces a single-space whitespace character. A backslash followed by a comma gives a small space. A tilde, ~, gives a "non-breaking" single-space whitespace character (meaning that LaTeX won't allow a line-break to occur over that space).

6. Spacing. A full-stop (period) is treated by LaTeX as a sentence-end, and LaTeX leaves extra space afterwards as a consequence. If you use a full-stop in an abbreviation, you should put a backslash afterwards to prevent this extra space:
   `e. g. this is an example. New sentence.` produces
   e. g. this is an example. New sentence. Compare with
   `e.\ g.\ this is an example. New sentence.` giving
   e. g. this is an example. New sentence.

7. Use math-mode, `$xxx$` for *all* formulae within the text (i.e., not within `\begin{equation}` `\end{equation}` pairs). This makes sure symbols such as $<$, $>$ and $-$ appear correctly in particular. For comparison, here is how $<$ $>$ $-$ appear when not in math-mode: ¡ ¿ - . No, that's

1

not a typo in my notes, that's how they appear. Note in particular that a hyphen – is not a minus-sign. This is a hyphen: - . This is a minus sign: −. Math-mode around full equations ensures consistent usage of type-face for variables, such as redshift z instead of z.

8. Math-mode is not the same as italics! The math-mode font in LaTeX is different from the italic font, plus math-mode treats white-space differently from text-mode, so don't use it as a short-cut for the `\em` or `\it` approaches to generating italics. Here is a comparison:
    `{\em This is correct italics}` gives *This is correct italics*.
    `$This is not italics at all$` gives $This is not italics at all$. In addition to the lack of whitespace note the different separations between characters, and the different type-face.

9. Hyphen, n-dash, m-dash, minus-sign. The single character – is used to generate all these char-acters. They are different textual constructions. A hyphen, the character with no special LaTeX markup, is used to hyphenate words as in Bottomley-Smythe. An n-dash, generated by `--` is a longer dash (nominally the width of an "n" character), and can be used as a colon to offset text – like so. An m-dash, `---` is longer still ("m" character width) — m-dashes can be used like this — in place of parentheses. A minus sign `$-$` is different again: $3 - 10 = -7$.

10. Angle-brackets for mean quantities are not the same as greater-than/less-than symbols.
    This is correct: `$\langle x \rangle$` gives $\langle x \rangle$.
    This is not: `$<x>$` gives $< x >$.

11. Different symbols used for approximately-equal-to. These include `$\sim$`: $\sim$, `$\approx$`: $\approx$, or `$\simeq$`: $\simeq$. While $\sim$ is commonly used now to mean $\approx$ (because there is no $\approx$ key on a keyboard, but the tilde key ˜ gives something like a $\sim$), many journals reserve $\sim$ to mean proportionality as opposed to approximately-equal-to (although of course `$\propto$`: $\propto$ should preferentially be used for this). The $\approx$ symbol should be used for approximately-equal-to in order to avoid ambiguity or imprecision.

12. Make sure to run LaTeX twice, in order for all references to be collected and appear correctly in the text. The first LaTeX run fills in the .aux file with references from the .tex file, the second run correctly propagates the .aux entries into the .dvi (or .pdf) output.

Common astronomy errors

1. When referring to magnitudes *always* use "brighter/fainter" to describe things. *Never* use "big-ger/smaller", "greater/less", "lower/higher" or similar constructions. Since magnitudes are nu-merically smaller for brighter objects, these latter constructions inevitably lead to ambiguity or confusion. Phrase your descriptions in terms of brightness and this will always remove such am-biguity.

2. A similar warning holds when discussing sensitivity and noise levels, or resolution. In particular, the words "higher" and "lower" can be misleading if you try to mix the concepts of sensitivity and noise in the same sentence. Terms such as "improved" or "better" (or their opposites) can be used to replace "higher" (or lower) to avoid ambiguity.

3. Infrared is one word. So is ultraviolet, redshift and extragalactic. They are not capitalised. They are not hyphenated.

4. When you compare positions of sources in multiple catalogues to identify common objects, this is termed "cross-matching," not "cross-correlation," which means something else entirely.

5. The symbol for the angular correlation function really is "w" not the greek omega "ω".

6. If you say the Galaxy (capitalised) you mean our Milky Way Galaxy. If you say galaxy (no capital) you mean any old generic galaxy. When referring to the Milky Way, you should capitalise the word Galaxy.

Figures and layout

1. Size of axis labels and numbering on figures. Make sure this is readable! Look at a collection of figures in papers and see what a typical size font is for figure labels/axes/keys, etc. Many common plotting tools, including IDL and Matlab, have stupidly small font sizes for axis numbering and labels by default. Get into the habit of specifying your font sizes explicitly in whatever plotting tools you use rather than relying on the default to give you anything sensible.

2. Multi-panel figures. Always label your panels (a), (b), (c), etc., and then refer to them as such in the caption and text. This avoids the clumsiness of referring to "top panel", "left panel", "top-left bottom-middle upside-down panel".

3. Refer to figures in the text using LaTeX such as `Figure˜\ref{great_fig}(a)` to give you the description of Figure 8(a), say. (The tilde here ensures no line-break between the word "Figure" and the number of the figure.) On this note, always use `\label{}` to give your figures a label and refer to them in the text with `\ref{}` accordingly, so figure numbering is always correct.

4. Put the `\label{}` command within the `\caption{}` braces to ensure the figure label is the correct number. E.g., `\caption{This is a figure caption. \label{great_fig}}`. Incorrect positioning of `\label{}` can lead to section numbers being used instead of figure numbers.

5. You can help LaTeX to position figures using the options `[h]`, `[t]`, `[b]`, or combinations of these, after the `\begin{figure}` command:
   `\begin{figure}[ht]`
   This requests LaTeX to position the figure at the top (t) or bottom (b) of a page, or "here" (h), meaning (to the best of its ability) to position it "here" relative to the manuscript in the LaTeX file around it. If you find figures always being pushed toward the end of your document, move the figure block in the LaTeX file further up. LaTeX won't position a figure earlier than it appears in your file, respective to the text.

6. You can modify or fix page layout with a LaTeX preamble command such as `\topmargin=-5 mm` (or equivalently `\renewcommand{\topmargin}{-5mm}`).

Text and English

1. Use a spell-checker. There are many. They can even be taught English English as opposed to American English. Even in LaTeX you can survive the spell-checker pulling you up on the markup if required. There are tools that spellcheck while avoiding markup (such as Excalibur with TexShop for the Mac), so look around for one that suits your needs.

2. In text where you use (a), (b), (c), or (i), (ii), (iii) to manually enumerate a list of items, use both brackets, not the ugly a), b), c), construction.

3. Firstly. Don't use "Firstly, Secondly, Thirdly", instead just use "First, Second, Third". It's briefer, punchier and clearer to read. Don't use "First[ly]" at all if there is no "Second" etc., following.

4. Avoid starting sentences with a numeral. Either write out the number in words, or rephrase to ensure the number comes later in the sentence.

5. Don't capitalise the words comprising an acronym. The acronym needs to be capitalised, but not the sequence of words. E.g., use "star formation rate (SFR)," **not** "Star Formation Rate (SFR)."

6. Don't overuse colons or semi-colons (: or ;). They are useful pieces of punctuation, used appropriately and in moderation, but you can almost always replace them by full-stops or commas respectively. This helps to shorten sentences and keep your discussion clear, as well.

7. Don't use a dash in place of a colon or semi-colon. This is informal usage, often seen in email, but not appropriate for a publication. You can almost always replace by a full-stop and shorter, clearer sentences. If you absolutely must, use a colon instead (such as when introducing an itemised list).

8. Equations. These need to be punctuated as if you were reading them as part of a sentence. They need to be followed by a full-stop or comma as appropriate, with following text capitalised or not as appropriate. The following text should not start a new paragraph (i.e., should not be indented) if it is in the middle of a sentence (beware leaving a blank line following the `\end{equation}`). An equation needs to be introduced as part of the sentence too. Here are some examples of correct punctuation of equations:
Everyone remembers the Pythagorean Theorem,

$$c^2 = a^2 + b^2, \tag{1}$$

where c is the length of the hypotenuse, but not everyone speaks Spanish.
Everyone remembers the Pythagorean Theorem:

$$c^2 = a^2 + b^2. \tag{2}$$

Here c is the length of the hypotenuse. Not everyone speaks Spanish.

9. However, Therefore, Moreover, Thus (and similar conjunctions). Do not use at the start of a sentence (with very rare exceptions). Do not use at all if your sentence reads perfectly well without them. While not authoritative, my preference is to avoid putting such conjunctions at the start of a sentence. People read the beginnings and ends of sentences, not the middle (just like papers, really). So put the things you want them to see there, not the joining words, which can go in the middle. Compare "However, our results contradict those of Amadeupname et al." with "Our results, however, contradict those of Amadeupname et al."

10. As such. This is an overused (and clunky) conjunction and can almost always be omitted. If you really need to link the previous phrase to the following one, try "accordingly," "consequently," "as a consequence," "correspondingly," "subsequently," "as a result," or just the simple "so," or even "therefore," "hence," or "thus," if you must.

11. The word "lead" is pronounced as "l-ee-d" (I am leading this expedition) or "l-e-d" (lead is a metallic element). The past tense of "lead" as in "expedition" is "led", *not* "lead" (which, as we have seen, is a metallic element).

12. Very. You can almost always omit every instance of "very" without any impact.

13. Formally $\neq$ formerly. "Formerly" means before, earlier, previously, as in "That miscreant was formerly our Prime Minister." "Formally" relates to being formal or proper, as in "Until you have a PhD you can't be formally recognised as a member of this exclusive club."

14. Be quantitative. Rather than saying "many" or "a large fraction" or "a large number" of some sample or population, explicitly say *how* many. This lends much more authority to your statement.

15. Plurals of Latin, Greek, or other-language derived words.
    "Data" is plural, "datum" is the singular. Where you talk about your data, you should refer to "these data" rather than "this data", for example.
    "Criteria" is plural, "criterion" is the singular.
    "Spectra" is plural, "spectrum" is the singular.

16. Then ≠ than. If x then y. I would rather eat broken glass than go out with you.

17. Affect ≠ effect. Detection of the Sunyaev-Zeldovich effect can affect my chances of getting a PhD. Your puny attempts to affect my mighty brain are ineffectual.

18. Loose ≠ lose. Lose is what you do to your phone, your glasses, and your wallet. Loose describes your belt, or your change. You can loose the hounds, but you can't loose your data. Maybe remember with the lame, and nonsensical: *Loose* needs to *lose* an "o". Or make up something better (and tell me so I can put it here).

19. "Classed." This is not the word to use when you actually mean "classified."

20. A compliment (with an "i") is something nice you say to someone. Also, the drinks at a fancy event might be complimentary. My data set, however, is complementary (with an "e") to yours. These words are not interchangeable. If you think you have a complimentary data set, it must be saying nice things about you (or maybe it was a free promotional item, and you know how much they're worth).

21. If you have a parenthetical phrase (such as this one), the punctuation (comma, fullstop) comes outside it (as here). (If, on the other hand, your entire sentence is parenthetical, like this one, then the closing full-stop comes inside the parentheses.)

22. "Sought" is the correct word if you're after the past tense of "seek". The word "sort" means to order.

23. The word "it's" means "it is". There is no apostrophe in the possessive "its". Take a random shopkeeper sign, for example. It's appalling that its grammar is so bad.

24. Plot. High school students plot graphs, and evil masterminds plot to take over the world. Respectable scientists, on the other hand, show figures, illustrate results, present correlations (or a lack thereof), and generally have a much broader vocabulary and list of synonyms to hand than merely doing a lot of plotting (which makes them sound like an evil mastermind...)

Additional reading
Ned Taylor pointed me to this URL, which has suggestions for improving your manuscript writing:
`http://matt.might.net/articles/`
`shell-scripts-for-passive-voice-weasel-words-duplicates/`
This is a good sequence of advice to follow as well, and comes with handy code implementations.