

Eww...

Let's pause to enjoy this attempt at writing (actual paper!):

Definition 5 Suppose $B = \{b_1, b_2, \dots, b_n\}$ is a set of backup tasks, then $B_i^+ = \{y \mid b_i \in B \text{ and } y \text{ is used to be denote a set of successor tasks of } b_i\}$.

(Source: Deng, Zheng, Wang, Hu, *Research and implementation of management software for SAN based on clustering technology*, ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry, 2004.
dl.acm.org/citation.cfm?id=1044689)

Realistic English expectations

*The basic problem in writing mathematics is the same as in writing biology, writing a novel, or writing directions for assembling a harp-sichord: **the problem is to communicate an idea**. To do so, and to do it clearly, you must have something to say, and you must have someone to say it to, you must organize what you want to say, and you must arrange it in the order you want it said in, you must write it, rewrite it, and re-rewrite it several times, and you must be willing to think hard about and work hard on mechanical details such as diction, notation, and punctuation. That's all there is to it.*

— Halmos, *How to write mathematics*, 2009.

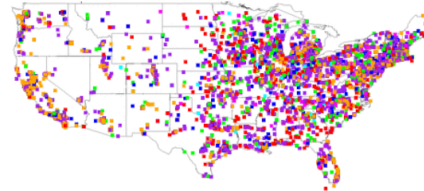
- ▶ **There is no such thing as perfection.** The best anyone (even a native English speaker) can achieve is reaching the “personal preference” level. What I mean...
 - ▶ There are some language matters which are **simply right or wrong**.
E.g. spelling “computer” as “comptuer” is simply plain wrong.
 - ▶ There are some language matters where **one view is favored over the other(s)**.
E.g. I would argue writing “the proposed method” is better than “our method” as it’s impartial and not exclusive.
 - ▶ There are some language matters where there are **multiple valid views**.
 - ▶ If you write X, some critics will say you should write Y. If you write Y, other critics will say you should write X.
- This is the “personal preference” level.
- E.g. Strunk & White say to write e.g. “Charles’s friend” instead of “Charles’ friend”. This is a matter of personal preference.
- However, whatever your preference is, **be consistent**. Don’t write both “Charles’s friend” and “Charles’ friend” in the same work.

“Dinner” and “supper”

96. What is the distinction between dinner and supper?

- a. supper is an evening meal while dinner is eaten earlier (lunch, for example) (7.79%)
- b. supper is an evening meal, dinner is the main meal (7.76%)
- c. dinner takes place in a more formal setting than supper (12.12%)
- d. there is no distinction; they both have the same meaning (34.56%)
- e. I do not use the term supper (33.14%)
- f. I don't use the term dinner (0.82%)
- g. other (3.83%)
(10661 respondents)

All Results



Source: Harvard Dialect Survey,
http://www4.uwm.edu/FLL/linguistics/dialect/staticmaps/q_96.html

(PS. I grew up saying “tea” instead of either of these.)

If you get to the “personal preference” level, your writing is fine. **Aim for this!**

This is how I intend to mark your assignments and exams.

Other points:

- ▶ **Languages change over time**, particularly because of the Internet. (What was once a mistake, might now be considered correct.)
- ▶ **British vs. American English** (e.g., standardised vs. standardized). Most Chinese people seem to be taught American English. Either is fine, *provided you are consistent*.
- ▶ **Who is your reader?** Probably a non-native-English speaker. Also probably a Master's or PhD student. Probably someone who doesn't want to read your paper. They probably won't read the whole paper.
- ▶ **It's okay to break the rules** when breaking the rules helps communication. (Don't overdo this, though.)

- ▶ As a native English speaker, I once had a reviewer explain that the English in my paper was poor. *If I cannot stop this happening to me, then you cannot stop this happening to you!!*

I guess they're a non-native English speaker, and had trouble understanding particularly complicated sentences.

Importantly, **this does not mean that I was right and they were wrong!!** (Wait, what?)

- ▶ This reviewer was a reader of my paper, a member of my “audience”.
- ▶ My purpose in writing a paper is to communicate my ideas to my audience. I did not succeed.
- ▶ I responded by simplifying some of the more complicated sentences.

If the reader does not understand, the author has failed to communicate!

Nowadays, scientific research is communicated through broken English.

- ▶ Sometimes people are unaware of differences of opinion.
 - ▶ I had a referee tell me to write “PLR(r, s, n)'s” instead of “PLR(r, s, n)s”, despite it being considered inappropriate at Oxford Dictionaries¹.
 - ▶ I had a referee write a whole paragraph about how they were confused about “nr” as an abbreviation for “number of” (instead of “no.”). It's simply a less-common alternative².
 - ▶ There's strong opinions about “less” vs. “fewer”, but it's complicated; see en.wikipedia.org/wiki/Fewer_vs._less.
 - ▶ Popularity of “gray” vs. “grey” has changed over time:

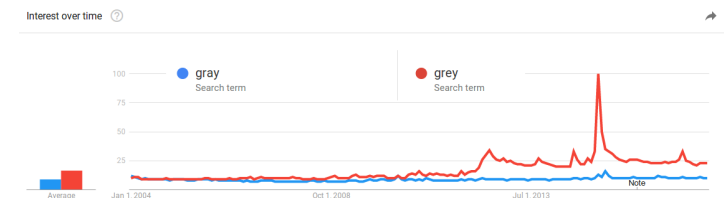


Image source: Google Trends.

¹<https://en.oxforddictionaries.com/punctuation/apostrophe>.

²<https://en.wiktionary.org/wiki/nr>

Take-home lessons

1. There's no such thing as perfection.
2. Technically correct writing is the minimum we should expect—we also need to write well.
3. We shouldn't think of English grammar as a list of rules to obey.
4. For almost everything I tell you, someone will disagree with me.

Other points:

4. Sometimes I see student papers which need rewriting from start to finish. Correcting sentence-level grammar will not fix this problem!!
5. What you consider “the best approach” will change over time.
6. Your coauthors will edit your paper too, and add their own English-language errors.
7. Even native English speakers are human and simply make mistakes.

Further reading

Further reading

What references benefit you the most will depend on your individual strengths and weaknesses. But here's some...

- ▶ Strunk and White, *The Elements of Style* (circa 1920), is probably the most famous English-language style guide.

It is still an important reference, but it's sometimes too strict or outdated. Some of the material is unnecessary since LaTeX does it for you (e.g. spacing). It's available online by searching for the title.
- ▶ *The Chicago Manual of Style* (originally published 1906) is the other famous English-language style guide. It's huge!

- ▶ Prof. Douglas West, a mathematician at Zhejiang Normal University (in Jinhua, China), has a webpage *The Grammar According to West* faculty.math.illinois.edu/~west/grammar.html.

It contains a list of advice for English in mathematical contexts. Most of this is useful, particularly when combining English with mathematics in sentences.

- ▶ Prof. Terrence Tao gives advice on writing papers terrytao.wordpress.com/advice-on-writing-papers/.
He talks about writing mathematics papers, but most of the advice applies to computer science papers also.
- ▶ Knuth, Larrabee, and Roberts have course notes *Mathematical Writing*: jmlr.org/reviewing-papers/knuth_mathematical_writing.pdf
- ▶ There's also question and answer sites:
 - ▶ ell.stackexchange.com (English Language Learners), and
 - ▶ english.stackexchange.com which is higher-level; there are many questions where the answer is “there's a difference of opinion”.

Bear in mind: *No book will solve your context-specific English-language problems.*

Here's a snippet from a paper by a Nankai PhD graduate:

During data deduplication, on-disk fingerprint lookups lead to high disk traffic, resulting in a bottleneck. In this paper, we propose a "lazy" data deduplication method which buffers incoming fingerprints and performs on-disk lookups in batches, aiming to reduce the disk bottleneck. ...

J. Ma, et al., Lazy Exact Deduplication. MSST 2016.

No book can tell you:

- ▶ What the audience at the MSST 2016 conference knows.
- ▶ What is considered standard terminology for this specific topic ("deduplication", "fingerprints", "on-disk lookups", etc.).
- ▶ Which parts of your paper should be omitted (repetition, too wordy, insufficiently important, etc.).
- ▶ How important is, say, a mathematical model compared to experimental results in this particular field.
- ▶ And many more things...

These all affect what to write and how to write it.

Also bear in mind: *References contradict one another as to what is "best".*

How to Write Like a Scientist by Adam Ruben³, Science 2012, satirically describes several well-established writing rules that scientists blindly follow. For example, he writes:

Using the first person ["I" instead of (academic) "we"] in your writing humanizes your work. If possible, therefore, you should avoid using the first person in your writing. Science succeeds in spite of human beings, not because of us, so you want to make it look like your results magically discovered themselves.

Ruben mocks standard practice: the academic "we", including the author(s) and the reader(s).

In fact, modern scientific thought is changing: more and more researchers are arguing that the experimenter is part of the experiment, and that "I" is preferable to "we" because it is more honest.

Take-home lesson: **Don't expect a book to teach you how to write your scientific paper.** At best, you might get some good general advice. You may also get bad advice.

尽信书不如无书

Jìn xìn shū bù rú wú shū

³<http://www.sciencemag.org/careers/2012/03/how-write-scientist>

Be concise!

While there's no "how do I write my paper" manual, one particular rule is widely accepted...

Strunk & White:

Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or that he avoid all detail and treat his subjects only in outline, but that every word tell.

Basically...

- ▶ No unnecessary words. No unnecessary sentences. **Don't waste the reader's time.**
- ▶ Every word should have a purpose...
 - ▶ We must go into detail sometimes, e.g., to precisely explain the experimental setup.
 - ▶ We also may prefer a longer explanation because a minimalist explanation is too hard to understand.

This is good, general advice; it has stood the test of time.

How to improve the writing in your specific paper

I don't know... just (a) do the best you can, and (b) when you find there's a better way, do that instead.

General advice (I've seen repeatedly):

Trial and error. Try several ways of writing something; include the best in your paper. Over time, this will become faster as you recognize (a) what did not work previously (and *why* they were poor), and (b) what worked well (and *why* it worked well).

How do other authors explain something similar? Identifying how other authors have written something (and comparing it to your own writing) helps context-specific writing. However:

- ▶ Do not just copy/paste what they write. (a) It's plagiarism, and (b) their paper has a different context.
- ▶ Think about what they wrote and how it works. Then put their paper aside and write your sentence using that knowledge.

(It helps to take a mental note whenever you read a sentence that makes you think "gee, that was explained well.")

How to improve the writing in your specific paper (cont.)

Read it out loud as if you're talking to someone. If it sounds wrong, it probably is.

(This has the added benefit of being good practice for when you present your work at a conference.)

Think "how would I explain this to a person sitting across from me?" Then write what you would say.

Get others to proofread your paper. (Other students perhaps.) Can they understand it? If not, it's not because they're unintelligent, it's because you haven't explained it well enough—it needs editing!

They need to give you blunt feedback, as if they were reviewing the paper and do not know you personally.

How to improve the writing in your specific paper (cont.)

It takes effort and time. It's normal for me to spend an hour or more on a single paragraph, perfecting it. Your name is attached to every error in your publications—it is your error, for life.

How many times does a paper need proofreading?

I'm not sure, but the number is not 0.

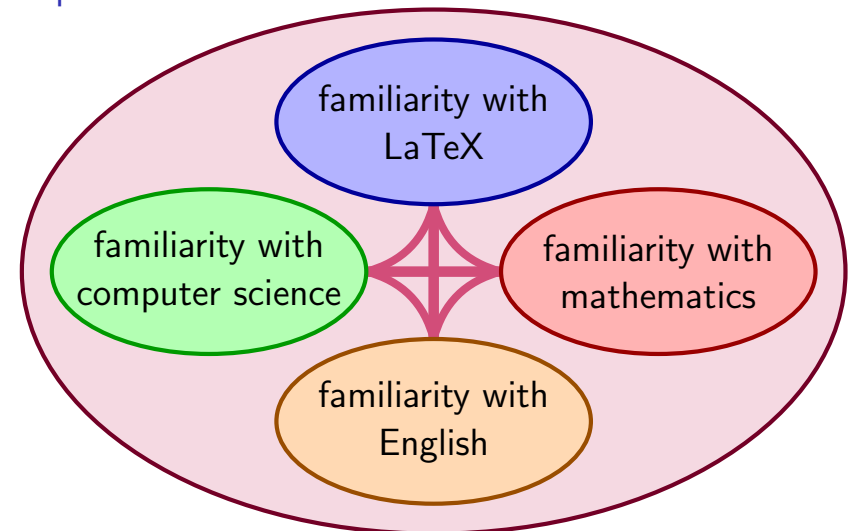
If you are still finding typographical or other errors in the paper while reading it, or if you are still adding results and commentary to the paper, then it is not yet time to submit the paper to a journal.

— Terrence Tao, *Submit a final draft, not a first draft*⁴, c. 2007.

⁴<https://terrytao.wordpress.com/advice-on-writing-papers/submit-a-final-draft-not-a-first-draft/>

English and LaTeX in computer science papers

Required skills



To write a computer science paper, you need all these skills. Moreover, you need to be familiar with the interactions between them. (And know how to Google stuff when you get stuck.)

Mathematics (inline equations)

We write English-language scientific papers using sentences. This does not change with the presence of mathematics.

Inline mathematics examples:

If there is a red edge between vertices i and j , then $\phi_2(i) = \phi_2(j)$ and $\phi_3(i) = \phi_3(j)$.

In this way, we ensure $\text{GS}(d') \leq \text{GS}(d)$ whenever $d' \geq d$.

Since L is diagonally cyclic, the symbols which occur in row i form the multiset $\{\theta(j) + i\}_{j=0}^{n-1}$.

When we read these sentences, we mentally replace the non-trivial mathematics with a no-context placeholder like “foo” or “blah”. This serves as a placeholder, since actually reading the mathematics would obstruct the sentence structure.

What's wrong with these sentences?

Let $f(x) = x^2, x \in \mathbb{R}$.

- ▶ Inline mathematics only separated by a comma is ambiguous.
- ▶ Better is: Define $f(x) = x^2$ for all $x \in \mathbb{R}$.
- ▶ We might make an exception for simple lists like “Let a , b , and c respectively denote ...” if reorganizing the sentence would be clumsy.

Define $S_n := \{1, 2, \dots, n\}$.

- ▶ We say “Define” once in words and once with the notation “:=”.
- ▶ It should be: Define $S_n = \{1, 2, \dots, n\}$.

Let n be a whole number, and let $m = 2n + 1$.

- ▶ The definition of “whole number” varies from place to place. (Like “dinner” and “supper”.)
- ▶ Better is: Let $n \in \{0, 1, \dots\}$, and let $m = 2n + 1$.

What's wrong with these sentences? (cont.)

For all $0 \leq k \leq n$, define $b_k = \binom{n}{k}$.

- ▶ When you read it, it begins “For all 0 less than or equal to k ...”.
(How many 0's are there?)
- ▶ Better is: For all $k \in \{0, \dots, n\}$, define $b_k = \binom{n}{k}$.
- ▶ Alternatively: For all k such that $0 \leq k \leq n$, define $b_k = \binom{n}{k}$.

Define $L_{ij} = i + j \bmod n$ whenever $1 \leq i, j \leq n$.

- ▶ Unclear if “ $1 \leq i, j \leq n$ ” means “ $1 \leq i$ and $j \leq n$ ” or “ $1 \leq i \leq n$ and $1 \leq j \leq n$ ”.
- ▶ Better is: Define $L_{ij} = i + j \bmod n$ whenever $i, j \in \{1, \dots, n\}$.

What's wrong with these sentences? (cont.)

M is an $m \times m$ submatrix with determinant 1.

- ▶ We don't start sentences with mathematics. (Some journal editors will correct this.)
- ▶ We could add in useless words: Note that M is an $m \times m$ submatrix with determinant 1. (Sometimes this is necessary.)
- ▶ In this case it's better to restructure the sentence: The $m \times m$ submatrix M has determinant 1.
- ▶ (In fact, saying “the vector \vec{v} ”, “the posting list ℓ ”, etc., is generally a helpful reminder for the reader.)

What's wrong with these sentences? (cont.)

Define *number_of_blocks* to be the number of blocks.

- ▶ 1. It's equal to $n \times u \times m \times b \times \dots$
(This problem could be resolved by using e.g. $\mathrm{\{...\}}$, which displays as *number_of_blocks*.)
- ▶ 2. This variable name is long... and being long does not help the reader.
- ▶ Better is: Define N_{blocks} to be the number of blocks.
Use either $\text{\$N_{\text{blocks}}\$}$ (which displays as N_{blocks}) or $\text{\$N_{\mathrm{blocks}}\$}$ (which displays as N_{blocks}) in LaTeX. These appear (and behave) slightly differently—pick one and be consistent.

However, if a variable is used very few times (once or maybe twice), it's often better not to give it notation.

We can write things like

$$\text{no. blocks} = \frac{\text{no. threads}}{\text{no. threads per block}},$$

instead of making the reader learn three once-off definitions.

What's wrong with these sentences? (cont.)

The probability of flipping exactly 3 heads from 5 coin tosses is $\frac{\binom{5}{3}}{2^5}$.

- ▶ The mathematics is too big to be an inline equation.
- ▶ Should be either (a) a displayed equation, or (b) rewritten $\binom{5}{3}/2^5$.

Let B be a $n \times n$ block.

- ▶ Should be “an $n \times n$ block” since we pronounce the letter “n” as if it starts with with a vowel (i.e., we say “en” which starts with “e”).

(See: https://en.wikipedia.org/wiki/English_alphabet.)

What's wrong with these sentences? (cont.)

If $f : \mathbb{R} \rightarrow \mathbb{R}$ is the function defined by $f(x) = x^2$, then $\forall x \in \mathbb{R}$ we have $f(x) \geq 0$.

- ▶ Notation such as \forall ($\text{\$forall\$}$), \exists ($\text{\$exists\$}$), \therefore ($\text{\$therefore\$}$), \because ($\text{\$because\$}$), \implies ($\text{\$implies\$}$), and so on, are useful for:
 1. When formally presenting propositional logic⁵.
 2. As a **blackboard shortcut** (when writing out these words in full is too slow).
- ▶ We should replace “ $\forall x \in \mathbb{R}$ ” with words “for all $x \in \mathbb{R}$ ”.
(We communicate with humans using words.)

Define $f(i, j) = (i * j) \% n$ for all $i, j \in \{1, \dots, n\}$.

- ▶ The symbols asterisk (*) and percentage (%) are misused.
- ▶ Should be: $f(i, j) = ij \bmod n$ (typeset $\text{\$f(i,j)=ij \mod n\$}$) or $f(i, j) = ij \bmod n$ (typeset $\text{\$f(i,j)=ij \text{ mod } n\$}$).

⁵https://en.wikipedia.org/wiki/Propositional_calculus

Take home lesson

- ▶ When writing technical English-language sentences, mathematics and English interact with one another in a variety of ways.
- ▶ Thus, to write well, we need to:
 - ▶ manipulate equations and rename variables, and
 - ▶ ensure the English and mathematics do not conflict.
- ▶ Your ability to write English-language computer science papers is affected by:
 - ▶ your familiarity with LaTeX, and
 - ▶ your familiarity with rephrasing mathematical expressions to suit what you're writing.

(This is why assignments are in LaTeX.)

Mathematics (displayed equations)

Displayed mathematics examples:

Consequently, (1) and (2) imply

$$S(d', q) \leq \text{GS}(d') \leq \text{GS}(d)$$

for all queries q and documents d and d' with $d \leq d'$.

The number of partial Latin rectangles that are isotopic to P , i.e., the size of the isotopism class containing P , is

$$\frac{r!s!n!}{\#\mathfrak{I}(P, P)}.$$

We are still in a sentence in a displayed equation, and we still need to use the correct punctuation.

Mathematics (align* equations)

An align* environment example:

Hence, by inclusion-exclusion,

$$\begin{aligned} |\mathcal{D}_U| &= |\mathcal{B}_U| - \left| \bigcup_{\substack{V \subseteq C_m \\ V \supsetneq U}} \mathcal{B}_V \right| \\ &= |\mathcal{B}_U| + \sum_{\substack{V \subseteq C_m \\ V \supsetneq U}} (-1)^{|V|-|U|} |\mathcal{B}_V| \\ &= \sum_{\substack{V \subseteq C_m \\ V \supsetneq U}} (-1)^{|V|-|U|} |\mathcal{B}_V|. \end{aligned}$$

We are still in a sentence...

Mathematics (arrays)

An array example:

For example, if $r = s = n = m = 3$ and $\mathbf{e} = ((1, 1, 1), (1, 2, 3), (1, 1, 1))$, then

$$M(\mathbf{e}) = \begin{array}{ccc} \{1, 1\} & \{3\} & \emptyset \\ \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset \end{array}.$$

We are still in a sentence... (although sometimes I make an exception here).

What I've shown is just a sample—there is no complete list.

We need to identify clashes between mathematics and English “on the fly”.

More than anything else:

- ▶ The primary goal is to communicate with the reader.
- ▶ Be succinct, but not if it hinders communication.
- ▶ Try to get a feel for how the reader would interpret your writing.



Image source: Well 2 xkcd.com/568