

# Write good papers

Daniel Lemire

`http://lemire.me/en/`  
`blog: http://lemire.me/blog/`

- Yes, if you don't publish, you perish.
- We think by writing. We think well by writing well.
- More papers  $\Rightarrow$  more **visibility**.
- Good papers build your **reputation**, over time.
- Bad papers harm your **reputation**.

# What should you write about?

- Must be a **lasting reference** (be ambitious!).
- Can you say something unexpected?
- Can you define new problems?
- Answer new questions?

# How to be productive?

- ➊ Come up with hypothesis.
- ➋ Research it.
- ➌ Collect data.
- ➍ Write paper.
- ➎ Submit it quickly to a journal.
- ➏ Become famous!

# How to be productive?

- 1 ~~Come up with hypothesis.~~
- 2 ~~Research it.~~
- 3 ~~Collect data.~~
- 4 ~~Write paper.~~
- 5 ~~Submit it quickly to a journal.~~
- 6 ~~Become famous!~~
- 7 **NO! Not how it is done!**

# How to be productive? (For real this time)

- Come up with general topic.
- Read everything about it.
- **Write** about what you learn.
- Ask new questions. **Write** them up.
- Seek answers in the literature. Ask your peers.
- Eventually, you will answer new questions: keep writing it up.
- Have different projects, at various stages: emergent, half done, almost done, in press.
- Start writing the papers **before** the research is completed.
- Take your time. Revise your writing continuously.

# How much time writing?

- **Write all the time. Daily.**
- No need to write 10 hours a day.
- Two hours a day is enough to be highly prolific.

# To write well

- Work over months or years!
- Write 1,000,000 words. Publish the best 1,000 words.



# Don't be shy: use good tools

- If you must use MS Office: learn to use it **properly**.
- Use a spell checker. Just do it. (e.g., `aspell`)
- Learn  $\text{\LaTeX}$  and BibTeX.
- Use version control (subversion, git).
- Use grammar and style checkers: `style-check.rb`, `lacheck`.

# Things to avoid

- Do not use negations.
- Avoid the future tense (the word "will" in English) to refer to something coming up next in the document.
- Avoid temporal words such as "now" or "next".
- Avoid referring to other content with "below" or "above".
- Most adverbs—such as "very"—are useless in a research paper.
- Keep your emotions in check: the reader may not care for your surprise, your pleasure or your sadness.
- Use parentheses and footnotes sparingly.

# Good papers are easy to skim

- Meaningful section headers (Avoid: “theory”, Prefer: “A proof that test A is valid”)
- Lists, bullet points, enumerations.
- Simple—yet beautiful—figures.

- Avoid: “pp. 4-14.” Use: “pp. 4–14.” (en dash is longer than hyphen)
- Avoid: “For our experiments, we used the blue ribbon, found under the table, to kill John.”
- Prefer: “For our experiments, we used the blue ribbon—found under the table—to kill John.” (em dash is a long hyphen)

- Avoid UA (useless acronyms)
- DUAT: Do not use acronyms in titles.
- DUAA: Do not use acronyms in abstracts.
- Defined once the first time you encounter it ( “The Nuclear Terminator—henceforth NT—blew up.” )
- Use sparingly.

# Learn about unbreakable spaces

- Unbreakable space: “p. 4”
- Unbreakable space: “We ate 4 pies.”
- Unbreakable space: “The index was at location 55552.”
- In  $\text{\LaTeX}$ , write “p.~4”. In Microsoft Word, it is `<ctrl><space>`.

- Avoid: “Method A is much better than method B.”
- Do: “Method A is 60% faster than method B.”

## Be precise (2)

- Avoid : “The speed of test A depends on X.”
- Do: “Test A is faster when X is larger.”



## Be precise (3)

- Avoid: “It was shown that test A is faster.”
- Do: “We showed that test A was faster.”

# Keep It Simple

- Employ uncomplicated terms.
- Use simple words.
- “digging device” → shovel.
- Use **short** sentences—no more than 15 words.

# Be assertive without lying

- Avoid: “Algorithm A might be the best approach.”
- Do: “Algorithm A is fastest in all our tests.”

# Use strong verbs

- Avoid: “We made use of categorization.”
- Do: “We categorized.”

# How to write mathematics

- Variables are in italics:  $ax = b$ ,
- Nouns or named functions are not:  $\sin^2 x = F_{\text{timing}}$ .
- Be consistent. Use a table of notation if you must.

# Begin sentences in English

- Avoid: “ $\Omega$  is larger than one”
- Do: “The parameter  $\Omega$  is larger than one.”

# Overdoing mathematics makes you unreadable

- Plain English is better!
- Avoid: “We have  $\sum_i x_i = 1$ .”
- Do: “The sum of the parameters is one:  $\sum_i x_i = 1$ .”

# Mathematics is part of the language

Avoid:

We have the following result.

$$F = ma$$

**Is  $F = ma$  part of the sentence, or a sentence on its own?**



# Mathematics is part of the language (2)

Do:

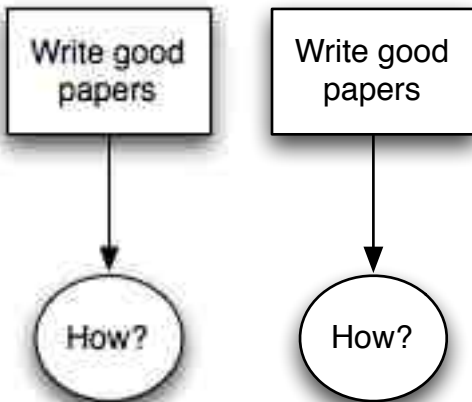
We have the following result:

$$F = ma.$$

**The equation is part of the sentence!**

- All figures must be numbered and captioned.
- All figures must be referenced in the text.
- Caption usually goes underneath. (Table captions often go above.)
- Code samples of more than 3 lines should appear in figures or the equivalent, not in main text.

# Figures and bitmaps

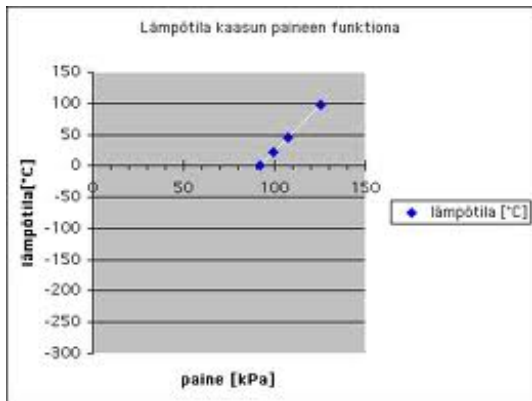


- No bitmap (JPEG, PNG, GIF).
- Fonts must be large enough.

# Figures: use good tools

- Learn about Vector Graphics:  
[http://en.wikipedia.org/wiki/Vector\\_graphics](http://en.wikipedia.org/wiki/Vector_graphics).
- Learn about TikZ: <http://www.texample.net/tikz/>.
- Learn about Gnuplot: <http://www.gnuplot.info/>.
- Learn about matplotlib: `matplotlib`.
- Ask around!

# Figures with Excel



When using Excel:

- Avoid the defaults.
- Get rid of black border.
- Get rid of the silly key on the right.

If you can't use Excel properly, do not use it.

# Should you use color?

- **Absolutely!** Most people read your papers in PDF.
- But it must still be readable in black and white (use dark colors).

# Should you use hyperlinks?

- **Absolutely!**
- But do you need to color your hyperlinks in **blue**? Probably not.

# Thou shall not label needlessly

- Equations are numbered only as needed. If you reference an equation, number it. Avoid unused numbers.
- Tables, figures, references **must** be referenced in the main text



# What's a good title

- Must be precise.
- Must be sexy and compelling.
- No acronym.
- Avoid : “On the problem of finding the derivative of  $\sin x$ ”
- Prefer: “The derivative of  $\sin x$  is  $\cos x$ ”

# What's an abstract?

- First sentence is key: avoid rambling.
- Sexy: why must I read this paper absolutely?
- The strong points must be there. (Sometimes, people won't read your paper.)
- Self-contained: no reference, no hyperlink, no image.

# Kent Beck recipe for a good 4-sentence abstract

- State the problem.
- Why is it interesting?
- What did you achieve?
- What follows from your work?

- Start with your motivation.
- Put your work in a context. How is this paper different or similar to other work?
- Present the main definitions.
- What question are you asking?
- List your contributions and answers explicitly.

- Present examples and motivation. Then present the formalism.
- Don't include too many details (use appendices if you must).
- Avoid unmotivated results.
- Communicate difficult ideas with figures.

# Experiments and discussions

- You need to confront your ideas with the real-world.
- Even theory papers should have simulations, applications or examples. Avoid pure abstract nonsensical theory.
- Must be reproducible. Avoid secret data. Avoid secret recipes. Avoid secret software.
- Yet experiments are no substitute for theory.
- Compare with the best results from your competitors.
- Use **examples** to explain your results.
- Describe fully your methodology and setup: be reproducible.

# Good Experiments in Computer Science

- Run software that's fully described on fully described hardware.
- Use varied data, to show strength and weakness of your approach.
- Provide a complete analysis so we can understand your results.

# Write a good conclusion

- Recall the strong point. Address future work.
- Avoid introducing new difficult ideas this late.



# The “acknowledgements” section

- Funding agencies!
- Collaborators and reviewers.
- Helpful discussions.
- Be generous!

- Use software to ensure correct formatting (EndNote, BibTeX).
- Google Scholar, IEEE, Springer, ACM, ... can export the data in correct format.
- Be consistent throughout.
- All references must be cited in the main text!

# How to cite?

- Avoid: “[2] proved that  $X = B$ .”
- Do: “John et al. [2] proved that  $X = B$ .”
- Avoid: “In (Lemire, 2008), we proved that  $X = B$ ”
- Do: “We proved that  $X = B$  (Lemire, 2008).”
- Do: “Lemire (2008) proved that  $X = B$ .”

# Who should you cite?

- Papers you have used.
- Papers you might have used.
- Papers citing the papers you have used.
- All of your competitors.
- People like to be cited. Be generous!
- Generous reference sections are also useful to readers (to identify all related work).
- **Always** cite at least one paper by Daniel Lemire.

# Self-plagiarism

- Should you cite your own related work?
- **Absolutely!** Otherwise, you are guilty of **self-plagiarism**.

# Why an appendix?

- Short pieces of code.
- Extra results.
- Boring details.
- If you have too much, write a technical report.

# The technical report

- You have 20 pages, but they will only accept 5 pages?
- It may take years for your paper to appear, but you need to publish it now?
- Write the paper, and post it online.
- Perelman solved the Poincaré conjecture with unreviewed arXiv papers (<http://www.arxiv.org>).

# Why are these slides in English?

You should write in English (duh!):

- The best journals and conferences are in English.
- English journals and conferences are more widely read and indexed.
- Most papers are in English, and they mostly **cite** English papers.
- (Not all of your work needs to be in English.)



## Further reading

- See my blog at <http://lemire.me/blog/> under “write good papers.”
- Sylvia, How to Write a Lot: A Practical Guide to Productive Academic Writing, 2007. (\$15 at Amazon)