ACADEMIC SKILLS IN COMPUTER SCIENCE

Lecture 7: Writing Research Reports (3)

Markus Krötzsch

Knowledge-Based Systems

TU Dresden, 21st May 2019

Review: Writing well

Cut away deadwood!

Support your claims!

Don't be a weasel!

Avoid ambiguous and vague formulations!

Review: writing (not) good

Avoid redundancies	"RAM memory"	"RAM"
	"objectives that programmers intend to achieve"	"objectives of programmers"
Prefer simple words	"utilise"	"use"
	"a plethora of"	"many different"
	"a number of"	"several" or "many"
Use strong verbs	"provides a rationale for"	"justifies" or "explains"
	"makes available"	"provides" or "offers"
Drop empty qualifiers	"very related to our work"	"related to our work"
	"remained broadly unexplored"	"remained unexplored"
No slang & colloquialisms	"I bagged my M.Sc. degree"	"I acquired an M.Sc."
No over-blown claims	"our work has amassed impact and community-wide adoption"	(three known uses; two in own projects)
Use active voice	"two cases were considered in our evaluation"	"we evaluated two cases"

Goals for today

Learning goals of this lecture:

(1) Become aware of further common issues in English writing

(2) Learn about essential conventions regarding punctuation

Tense

> Research reports use mostly simple present and simple past tense

- Simple present for general truths: "SAT is NP-complete"; "Python is widely used."
- Simple present for atemporal facts: "In Section 4, we explain our approach." (not: "we will explain"); "This paper presents . . . "; "This observation suggests that . . . "; "By Lemma 3, it follows that . . . "; and "In their seminal paper, Feder and Vardi [1998] define MMSNP and state their famous dichotomy conjecture"

Tense

> Research reports use mostly simple present and simple past tense

- Simple present for general truths: "SAT is NP-complete"; "Python is widely used."
- Simple present for atemporal facts: "In Section 4, we explain our approach." (not: "we will explain"); "This paper presents . . . "; "This observation suggests that . . . "; "By Lemma 3, it follows that . . . "; and "In their seminal paper, Feder and Vardi [1998] define MMSNP and state their famous dichotomy conjecture"
- Simple past for prior steps and observations: "We ran the algorithm and measured total runtime"; "The modified system showed performance improvements"; "We failed to establish a tighter bound for the runtime of this algorithm"; "Feder and Vardi first stated their famous dichotomy conjecture in 1993"

Tense

> Research reports use mostly simple present and simple past tense

- Simple present for general truths: "SAT is NP-complete"; "Python is widely used."
- Simple present for atemporal facts: "In Section 4, we explain our approach." (not: "we will explain"); "This paper presents . . . "; "This observation suggests that . . . "; "By Lemma 3, it follows that . . . "; and "In their seminal paper, Feder and Vardi [1998] define MMSNP and state their famous dichotomy conjecture"
- Simple past for prior steps and observations: "We ran the algorithm and measured total runtime"; "The modified system showed performance improvements"; "We failed to establish a tighter bound for the runtime of this algorithm"; "Feder and Vardi first stated their famous dichotomy conjecture in 1993"
- Present perfect or simple past for earlier content: "So far in this book, we have focused on . . . " or "In Chapter 1, we introduced . . . " (mostly in longer texts)

Tense

- Simple present for general truths: "SAT is NP-complete"; "Python is widely used."
- Simple present for atemporal facts: "In Section 4, we explain our approach." (not: "we will explain"); "This paper presents . . . "; "This observation suggests that . . . "; "By Lemma 3, it follows that . . . "; and "In their seminal paper, Feder and Vardi [1998] define MMSNP and state their famous dichotomy conjecture"
- Simple past for prior steps and observations: "We ran the algorithm and measured total runtime"; "The modified system showed performance improvements"; "We failed to establish a tighter bound for the runtime of this algorithm"; "Feder and Vardi first stated their famous dichotomy conjecture in 1993"
- Present perfect or simple past for earlier content: "So far in this book, we have focused on . . . " or "In Chapter 1, we introduced . . . " (mostly in longer texts)
- Future for next steps: "We will investigate other applications for this algorithm in future work."

Tense

> Research reports use mostly simple present and simple past tense

- Simple present for general truths: "SAT is NP-complete"; "Python is widely used."
- Simple present for atemporal facts: "In Section 4, we explain our approach." (not: "we will explain"); "This paper presents . . . "; "This observation suggests that . . . "; "By Lemma 3, it follows that . . . "; and "In their seminal paper, Feder and Vardi [1998] define MMSNP and state their famous dichotomy conjecture"
- Simple past for prior steps and observations: "We ran the algorithm and measured total runtime"; "The modified system showed performance improvements"; "We failed to establish a tighter bound for the runtime of this algorithm"; "Feder and Vardi first stated their famous dichotomy conjecture in 1993"
- Present perfect or simple past for earlier content: "So far in this book, we have focused on . . . " or "In Chapter 1, we introduced . . . " (mostly in longer texts)
- Future for next steps: "We will investigate other applications for this algorithm in future work."
- Other tenses can be appropriate in special cases, e.g., "Some readers will prefer to skip this chapter" or "The Web will have doubled in size by the end of this decade."

Tense changes meaning

Note the difference:

> "Our implementation needed less memory than other systems."

⤳ we observed this (in our experiments)

> "Our implementation needs less memory than other systems."

⤳ this is a general truth

Parallel clauses

Parallel clauses, which follow the same structure, are used in all lists.

Example: The following sentence is badly violating this principle: "Approaches for improving performance include the use of faster CPUs, optimising existing algorithms, and to approximate solutions." Better versions:

- "... include the use of faster CPUs, the optimisation of existing algorithms, and the approximation of solutions."

- "... include using faster CPUs, optimising existing algorithms, and approximating solutions."

- "... are to use faster CPUs, to optimise existing algorithms, and to approximate solutions."

Lists and enumerations (whether inline or not) should always use parallel clauses.

(In contrast, sequences of parallel sentences in paragraphs are usually tiring.)

Format and Layout

Format and layout

There are several main aspects to formatting and layout:

- **Space**
- **Fonts**
- **Colour**
- **Arrangement**

General guidelines:
- Be consistent
- The goal is readability and clarity, not originality or attention grabbing
- Mind the medium (paper size? colour printing? etc.)

Space

- Margin sizes
- Spaces before and after various parts (headings, tables, figures, lines, etc.)
- Indentation (=shifting start of lines away from left margin) of paragraphs, lists, enumerations, etc.

Guidelines:
  - Follow spacing instructions where available
  - Be generous (empty space is the most precious resource in layout)
  - Follow conventions

Example: In English writing, there is usually no additional vertical space between paragraphs, but their first line is indented. In German, it is the opposite.

Writing something in very large font does not necessarily make it stand out or appear stronger. The same is true for **text written in bold** or *italics*. Even huge font sizes do not really have much effect. ALL-CAPS ARE EVEN WORSE, SINCE THEY IMPAIR READABILITY. The same *applies* to using different *FONTS*. The strongest form of emphasis is

empty space

since it is the ONLY EMPHASIS that does not distract from the content but instead makes the reader **focus** on it. Space guides and *clarifies structure*. This effect is widely used in layout, in printed and online media alike. Logos often include much **white space** or require **wide margins** for this reason. Note that emphasising too much achieves the opposite effect.

Fonts

- Family (Times, Comic Sans, etc.)
- Shape and highlight (normal, italic, bold, . . .)
- Size

Guidelines:
- Use a single main body font for most text (preferably serif, e.g., Times)
- Maybe use one or two other fonts for special cases (headlines, code, . . .)
- Math uses fonts for distinguishing identifiers – respect these conventions
- Use few sizes, defined by text structure (headlines, body text, footnotes, . . .)

Inline emphasis

Emphasis of words in text has several purposes:

- Main use: to highlight new technical terms when first defined
- Less common: when writing about a word rather than about its meaning

> **Example:** A few years ago, examples of neologisms still included *neologism*.

- Rarely: to set apart foreign language phrases
- Never: to emphasise an important message

How to emphasise:

- set words in *italics* or **boldface**
- italics leads to calmer overall look (often preferred)
- boldface stands out on page level (sometimes useful, e.g., for textbooks)
- underlining is never used in print (exception: some artistically styled books)

Colour

Colour is rarely used much, since many research publications (also) appear as black-and-white prints

Guidelines:

- Colour is very powerful – use with care
- You may use colour even if a report is also printed in black and white, since the digital version can still benefit from it, but do not rely on it in this case
- Avoid colours that are not readable or plain ugly

Arrangement

- Number of text columns
- Placement of figures and tables
- Use of special layouts (e.g., text flowing around a figure)

Guidelines:
- This is usually partly determined by applicable formatting guidelines
- LaTeX has good default solutions for arranging content
- Use special layouts sparingly

Punctuation

Punctuation

Punctuation is the use of punctuation symbols, most commonly:

- Full stop .
- Comma ,
- Semicolon ;
- Colon :
- Hyphen -, en-dash –, and em-dash —
- Apostrophe ', quotes " ", and single quotes ' '
- Parentheses ( )

Rules apply, but often depend on context (medium, style guide, preference).

Full stops and semicolons

Full stops

- end sentences that do not already have a . in the end for other reasons
- follow after abbreviations (incompl. words, not acronyms)
  Note: only use standard abbreviations (e.g., **etc.** or **Fig.**)
  Note: sentences should never start with abbreviations
- are not used in titles and headings

Semicolons

- separate sentences like full stops, if a closer link is desired (but are not require a capital letter to follow)
- can be used instead of commas in list (useful if list items contain commas)
- should be used, but not too much

Examples:

Semicolons connect phrases that could each be sentences; that's their main use.

Figure 2 uses the same format as Fig. 1.

A similar conjecture was made by Smith et al.

Colons

Colons strongly connect parts of text

- Can express **as follows** or **for the following reason**

> **Example:** There are two essential resources: time and memory.

> **Example:** The system was excluded from our evaluation: it ran out of memory in all experiments.

- Is mostly used inside a sentence (even if followed by what could be a sentence)
  ⤳ words after **:** are usually lower case
- Can also be used in a clause that initiates several sentences or a quote
  ⤳ in these (less common) situations, words after **:** start upper case

> **Example:** As Abraham Lincoln once put it: "The problem with quotes found on the Internet is that they are often not true."

Hyphens

Hyphens are used to connect words to form a whole

- Connect words split across two lines (usually done by software, but needs review)
- Build words from parts

> **Examples:** "hash-based", "near-memory computing", "co-operate" (British)

- Overwrite right associativity of longer noun phrases

> **Example:** "American history teacher" vs. "American-history teacher"

Note that hyphenation depends on context and intended meaning

> **Example:** Up until to date, state-of-the-art software is defined to be up-to-date with the state of the art.

> **Example:** The headmaster recovered from his embarrassment almost as quickly as he had re-covered his backside.

Dashes

Dashes are used to separate parenthetical remarks from the rest of the sentence

Examples:

- Many uses of dashes – including this one – could also use commas instead.
- If a sentence uses a comma – as in this example –, then the dash does not replace it.
- There is no closing dash at the end of a sentence – as in this example.

Dashes can be written in two ways:

- The shorter en-dash with spaces is more common in British spelling – and also in other European languages.
- The longer em-dash without spaces is common in American spelling—and is also preferred by some other style guides.

Never use em-dashes with spaces (empty gap visually too large).

Whichever variant you pick, be consistent.

Note: en-dashes can also be used to indicate ranges, as in "page 3–12"; no spaces are used in this case

Commas

Commas have many uses in English (non-exhaustive list)

- Items in a list or enumeration: "I like apples, mangos, and strawberries"

- Coordinated (equal-rank) adjectives: "the sweet, fragrant, mouth-watering scent of mangos"

  But only if they could swap places, e.g., no comma in "fast stochastic gradient descent" or "useless optimal solution"

- Introductory phrases of sentences:

  "However, air-shipped mangos are bad for the environment."

  "Considering the carbon footprint of his purchase, he bought fresh strawberries instead."

  Some such commas are optional: "In Section 4$_{(,)}$ we evaluate our approach."

- Parenthetical remarks and clarifying insertions:

  "The fastest algorithm, Quicksort, ran in under 45 seconds."

  "The fastest algorithm, which we therefore chose for our software, was Quicksort."

  "The fastest algorithm, as determined by averaging ten runs, was Quicksort."

  This comma can sometimes be optional: "And is was memory-efficient$_{(,)}$ too."

- Independent clauses joined by conjunctions

  "We met only twice, but/and/so/... we did not have dinner together."

  But not for compound verbs/nouns (i.e., lists of two): "We met twice and had dinner." and "We had dinner that night and breakfast next morning."

- Out-of-order dates: "May 21, 2019" (but: "21 May 2019")

English distinguishes defining and non-defining relative clauses.

That and which

English distinguishes defining and non-defining relative clauses.

Defining: (restrictive)

"We recommend the hash-based algorithm that was the fastest in all experiments."

- Implies that several hash-based algorithms were considered
- The specific algorithm that is recommended is defined by being the fastest

↝ deleting the relative clause would cause ambiguity

Non-Defining: (non-restrictive)

"We recommend the hash-based algorithm, which was the fastest in all experiments."

- Several algorithms have been considered, but only one was "the hash-based"
- This one happened to be the fastest

↝ deleting the relative clause would not cause ambiguity

Know the differences: no comma vs. comma; that can only be used in defining clauses
(which, who, where, why, whose, and whom can be used in either case)

E.g., i.e., and such as

Such as: use of commas based on defining vs. non-defining

- "Sorting algorithms, such as Quicksort and Bubblesort, are taught to undergraduates."
- "Algorithms such as Quicksort and Bubblesort are well-suited for teaching basic data structures."

e.g. and i.e.

- These abbreviations mean **for example** and **that is** (**in other words**), respectively
- They always start a parenthetical remark and must be preceded by a comma
- They may or may not be followed by a comma (your choice; be consistent)
- Another comma is needed to close the parenthesis (if not at end of sentence)
- Like all abbreviations, they should never start a sentence

Example: Sorting algorithms, e.g.(,) Quicksort and Bubblesort, are good for teaching basic algorithmic concepts.

The serial comma

> The final conjunction in a list may or may not be preceded by a comma.

The serial comma (a.k.a. Oxford comma):

- should be used (or omitted) consistently
- adds clarity

> **Example:** "I thank my parents, Madonna, and the Pope."

- appears slightly more formal
- is unrelated to the use of commas before **and** as a conjunction of independent clauses (where we could also use conjunctions like **but** or a full stop)

> **Example:** "We verified our algorithm, and we also ran experiments."

- is not used with two-item lists, esp. for correlative conjunctions (that come in pairs)

> **Example:** "Existing solutions are either fast or memory-efficient."

Apostrophes and quotation marks

Apostrophes are used in contractions, such as "can't"

- avoided in formal writing, but can be used in textbooks and informal reports
- most research report style guides discourage them, but some enforce them

Apostrophes also indicate possessives, as in "Tarjan's algorithm"

- Followed by **s** for singular nouns: "the algorithm's complexity"
- No **s** for plurals: "the users' comments"
- Optional **s** for singular words that end in **s**: "Markus' lecture" or "Markus's lecture"
- Never for possessive forms of pronouns: "its main reason" or "it was hers"

> The **Greengrocers' Apostrophe** is the erroneous use of apostrophes, especially in plural forms ending with s ("Fresh apple's!").

Quotation marks

Quotation marks are used for quotations and to mark literal uses of text

- Most styles prefer double quotation " ... " (shape in English: 66 ... 99), but one can also consistently use single quotes ' ... '
- The other style of quotes is used when nesting quoted strings inside quotes
- Different from straight quote " found on keyboards
- Many style guides move closing " after commas or full stops that follow immediately

> **Example:** "WWW," which is short for "World Wide Web," is one of the few abbreviations that take longer to speak than what they abbreviate.

The reason is aesthetics, but especially computer scientists often avoid this style to prevent (possibly disastrous) confusion when quoting source code, where punctuation often is significant

> **Example:** Delete the temporary files using the command "`rm -rf ./tmp`".

Parentheses

Quotation marks are used for parenthetical remarks

- Useful for brief insertions, especially in technical writing

> **Example:** Figure 1 compares Quicksort (left) with Bubblesort (right).

> **Example:** Artificial intelligence (AI) uses a variety of methods, such as Machine Learning (ML) and Knowledge Representation and Reasoning (KRR).
> Note: avoid the overuse of abbreviations!

- Should be avoided for longer phrases or sentences: use commas, dashes, or footnotes; or simply unravel your writing

Capitalisation

English uses upper case letters only rarely:

- sentence beginnings
- proper names (including identifiers such as "Theorem 3" or "Yannakakis Algorithm")
- most multi-letter abbreviations (depending on convention)
- titles and headings
  - Style 1: only capitalise start and words after ":"
  - Style 2: capitalise all words other than short common words such as "for", "on", or "and"

Some names are stylised to use unusual or distracting case

- Decide what is acceptable: readability weighs heavier than other people's extravagance (esp. lower case at sentence starts is irritating and puts style over content)
- Ignore any other special styles (colour, font, etc.) of names in regular text

Whitespace

Spaces are used between words, but also

- to separate parentheses from words on their outside

> **Example:** "...shown by Smith[42]." is wrong; "...shown by Smith [42]." would be correct.

- around (parenthetical) en-dashes
- after sentences (double spaces were used after full stops until the mid-20th century; this is mostly obsolete now)

Note: In English, there are no spaces before sentence marks or within long numbers

Summary

Research reports are mainly written in simple present and simple past

Empty space is the most important resource in document layout

Fonts, colour, and layout elements should be used in moderation

English punctuation differs from that in other languages in subtle ways

What's next?
- Typesetting with LaTeX
- Persuasive writing and argumentation
- Oral presentations