

# Writing a Scientific Article, part II

Vladimír Ulman, Igor Peterlík, Jan Obdržálek

FI MU

DUVOD, November, 2015

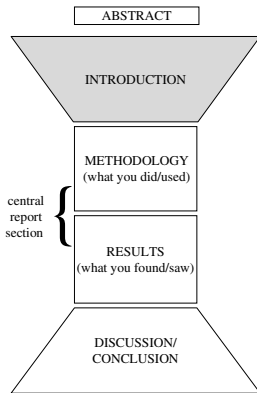


INVESTMENTS IN EDUCATION DEVELOPMENT

- English
- Algorithms
- Figures and charts
- Tables
- Bibliography
- Acknowledgement
- Short CVs
- Supplementary materials
- English and general typesetting (symbols, numbers, units)

- *Science writing is much easier than it looks.*
- find a good book, for example: *Science research writing for Non-native speakers of English* by Hilary Glasman-Deal (free)

1. READ THE TITLE  
and try to predict the type of information you expect to see
2. LOOK AT THE NAME OF THE AUTHOR  
What you know about the writer will help you predict and evaluate the content.
3. CHECK THE DATE  
and use it to help you assess the content.
4. READ THE ABSTRACT  
to find out what the researchers did and/or what they found
5. LOOK QUICKLY AT THE FIRST PARAGRAPH  
without trying to understand all the words.
6. LOOK QUICKLY AT THE FIRST SENTENCE OF EACH PARAGRAPH  
without trying to understand all the words
7. LOOK QUICKLY AT EACH FIGURE/TABLE AND READ ITS TITLE  
to try and find out what type of visual data is included
8. READ THE LAST PARAGRAPH  
especially if it has a subtitle like 'Summary' or 'Conclusion'



- Present simple
  - The most important and frequently used tense in scientific writing
  - Stating accepted facts and truths, permanent situation (introduction, s.o.t.a.)
  - Describing standard procedures (methods)
  - Reported findings are considered as general, another scientist will necessarily get the same results (results)
- Present continuous
  - Temporary situation (a change is expected) (intro)
- Past simple
  - What you did yourself in contrast to standard procedures (methods)
  - Indicates that findings are linked only to your research (results)
- Present perfect
  - The event reported is relevant to the actual situation (intro)
- “Simple future” (“will”)
  - To be avoided everywhere except in future work

- Correctness of the algorithm was demonstrated two years ago. However, little attention has been paid to other aspects such as...  
*vs.*

However, little attention was paid to other aspects such as...

- A global minimum was found using a genetic algorithm...  
*vs.*

A global minimum is found using a genetic algorithm...

- We found that the convergence of the iterative method strongly depended on...  
*vs.* We found that the convergence of the iterative methods strongly depends on...

- *The space between a full stop and the next capital letter is a dangerous space for you and your reader.*
- Therefore, use “connectors”
- They considerably improve fluidity and facilitate understanding
- Cause: since, as, because, due to, on account of
- Result: therefore, hence, consequently, so, which is why
- Contrast: however, whereas, but, while, by contrast, yet
- Unexpectedness: although, even though, despite, notwithstanding, regardless of, yet, nevertheless, nonetheless
- Addition: moreover, further, furthermore, also, secondly, in addition

# Scientific English: Talking about Results.

- *Results do not speak for themselves!*
- Sequence: order of things (extremely important in algorithms, but also on a “global level”). Useful words: first, second, then, afterwards, next, finally, prior to, subsequently, immediately. . .
- Frequency: how often an event or result occurred. Useful words: each time, always, without exception, constantly, repeatedly, frequently, sometimes, at times, occasionally, from time to time, rarely, seldom, hardly ever, barely ever, at no time, never. . .
- Quantity: great deal, important, considerable, significant, substantial, appreciable, plenty, reasonable, numerous, somewhat, particular, moderate, few, little, small, negligible, exceptional, infinitesimal, marginal
- Causality: relationships and connections between observations. Useful: cause, produce, originate, be due to, initiate, be related to, be linked to, contribute to, correlates with, be a factor in, results from, yield, stem from

- Usually using a lot of “modal verbs“:
- Ability/Capability: can, cannot, could, could have, could not. . .
- Possibility/Options: may, might, could, may have, might have. . .
- Probability/Belief/Expectation: should, ought to, should have, ought to have. . .
- Virtual certainty: must, have to, cannot, must have. . .
- Advice: should, ought to. . .
- Necessity: must, need to, have to. . .



# Choosing the Right Template

- Conferences usually provide templates, styles for contributions
  - $\LaTeX$  templates (LNCS, IEEE, Elsevier)
  - MS Word templates (mainly in non-CS domains)
- The style usually determines:
  - Page size, columns, headers, footers
  - Title, captions placement and numbering
  - Font type (text, captions)
  - Citation style
- The latex template often provides several “modes”
  - Mainly used in journal submissions (conference is usually in camera-ready template from the beginning due to the strict length)
  - Draft, review, final, one-column, two-column
- Some properties of the submission can be changeable (mainly in journals)
  - Number of pages; buying extra pages often possible
  - Black/white vs. color
  - *Discuss the financial resources with your supervisor.*

# Tables, Figures and Algorithms

- Should increase the comprehensibility of the text
- Every object must be referenced in the text
  - In  $\LaTeX$  use `label`, `ref`
  - The label is generated by the caption command, so `label` should be placed after `caption` to have correct references
  - Correct referencing:
    - "...in Table 1...",  $\LaTeX$ : `in Table~\ref{tab:label}`
    - "...in Fig. 2...",  $\LaTeX$ : `in Fig.~\ref{fig:label}`
- Placement on the page can be given by template or recommended by the conference/journal web.
  - In  $\LaTeX$  Here, Top, Bottom (forcing with with '!')
  - *Placement in the top can sometimes save some place (no top margin needed to separate from the text)*
- In journal submissions, tables, figures and algorithms may be added to the text separately (mainly in the first submission).
- In some journals, paper subsections (state of the art) can appear in highlighted boxes.

# Typesetting Algorithms

- Surprisingly not very often even in computer science papers
  - Maybe because of the idea *say enough but not too much...*?
  - Typesetting an algorithm should be considered carefully (and probably avoided in Word-like submissions)
  - A text description is sometimes sufficient (but increases *dense prose*)
  - Usually quite useful in thesis or longer reports
- Several  $\text{\LaTeX}$  packages exist for algorithmic environment
  - Algorithmic, algorithmicx, algorithm2e
  - [http://en.wikibooks.org/wiki/LaTeX/Algorithms\\_and\\_Pseudocode](http://en.wikibooks.org/wiki/LaTeX/Algorithms_and_Pseudocode)
- General rules
  - Show only essential and rather short functions/procedures
  - Rely on indentation (usually provided by the package)
  - Reduce (or avoid) `{ }` delimiting blocks
  - Reduce the occurrence of special symbols and long variable names

# Figures and Images

- *A (good) picture is worth a thousand words.*
- A figure: drawing, shape, visual representation
  - Typically a diagram (flowchart, function graph, data chart). . .
  - Can be very practical when giving the big picture
  - . . . or a graph (pie, bar, line charts)
  - Often necessary to present/illustrate the results
- An image: a picture
  - Taken by a camera, extracted from a video, screen shot
  - Can be very practical to illustrate the results or display “realia”
- Consider the requirements for the paper
  - Color, length, page size, style. . .
- A figure can significantly shorten the text (simplified explanations)
- Color limitation (b/w) can be an issue
  - Consider different plotting patterns (dotted lines, thickness, gray-scale etc.)
- Do not forget legend to interpret colors, lines and other objects in the figure

# Figures and Images

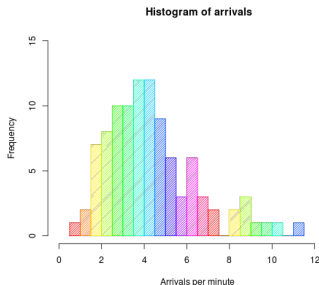
- Employ a *vector representation* as much as possible (figures)
- If pixel representation is unavoidable, keep a sufficient resolution (600 DPI = 600 pixels per 2.54 cm), convert to pixels if needed
- Avoid JPG if the paper is to be printed (not always possible...)
- Never use a figure (schemes, graphs, plots) that was saved in JPG in some phase of processing (blurred edges)
- Always consider readability
  - Contrast, color saturation, size of details (paper cannot be zoomed)
  - Font size if any text or symbols are present
- Organize the images and figures properly
  - Use `subfigure` package if allowed by the template
  - Distribute images over the article, prefer “top” placement
  - Page-wide placement is often even in two-column style (`figure*`, `table*`)
- Use short but self-explaining captions
  - Many readers inspect only title, abstract and figures
  - Always reference each figure in the text

- A powerful mean of data visualisation
  - Comparing data (e.g. gain, acceleration, scaling)
  - Showing an evolution, growth, development
  - Statistical evaluation
- Find your favourite application to create charts
  - GNUPlot, R, Matlab, Paraview
  - My basic requirements:
    - Allows for creating all standard types of charts
    - Allows for "scripting" the graph and re-generating (*my usual pipeline: create a script, render a graph, modify interactively, regenerate the script*)
    - Stores the rendered figure in vectorial format (PDF, PS, SVG)
- General rules:
  - Good readability, comprehensibility, contrast (b/w, colors)
  - Self-explaining short caption (title is not always necessary)
  - Correct annotation of all axes (physical units, amounts, descriptions)
  - Legend if needed

# Chart Types: Histogram, Pie Chart

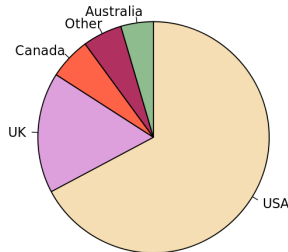
- Histogram:

- Graphical representation of distribution of data
- Show frequency of data (y-axis)
- A 3D version exists (with two range axis): questionable use in papers (mainly printed)



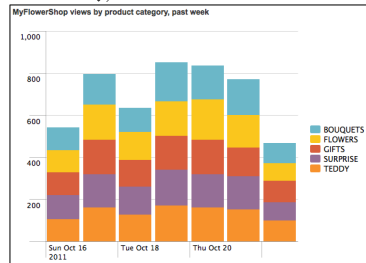
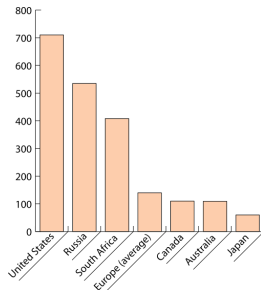
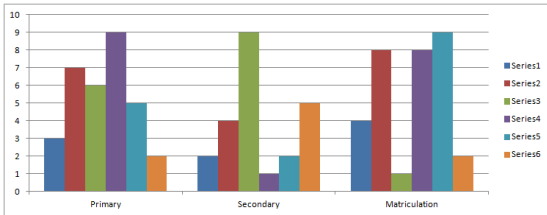
- Pie chart:

- Graphical illustration of proportion in data
- Assuming that data sums up 100%
- Often criticized: difficult comparison across different pie charts (often (mis-)used by media)



# Chart Types: Bar Chart

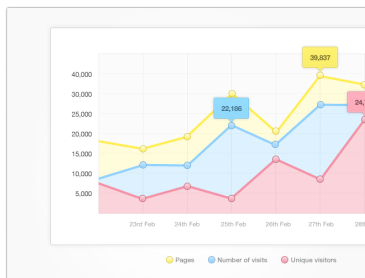
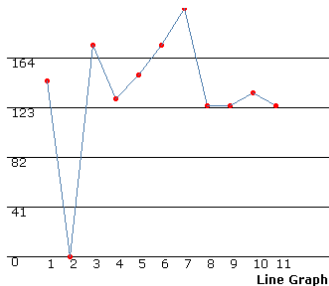
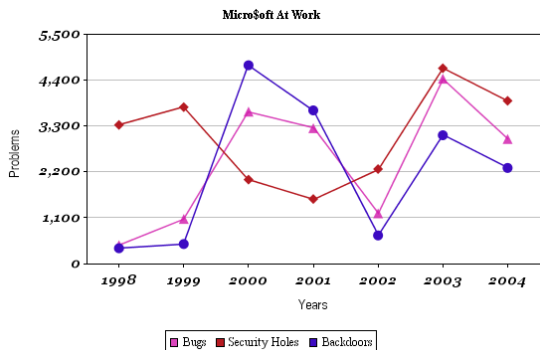
- Shows comparisons between categories (discrete range data)
- Can be plotted vertically and horizontally (important when shortening the paper)





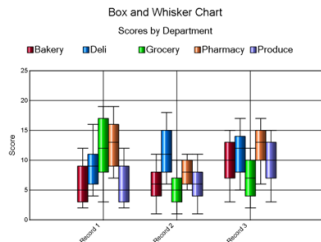
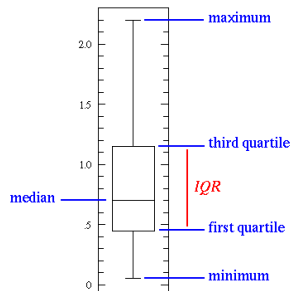
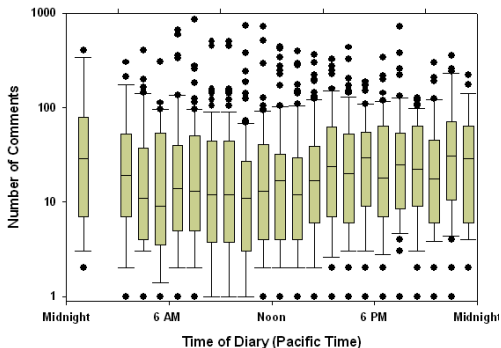
# Chart Types: Line Chart

- Series of data (depicted by markers), connected (or not) with a line
- Typical for time series or for example scaling
- Grid can be used in order to increase the readability



# Chart Types: Bar-and-whisker Plot

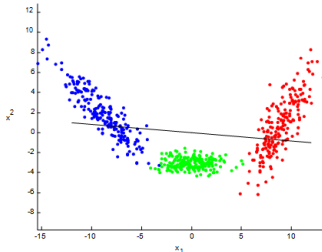
- Depicts groups of data using *quartiles* and extrema
  - Alternatively, whiskers depict  $\pm\sigma$  (std. deviation), percentiles
- Outliers can be shown (point markers)



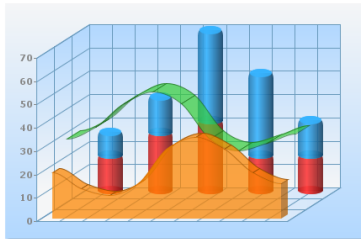
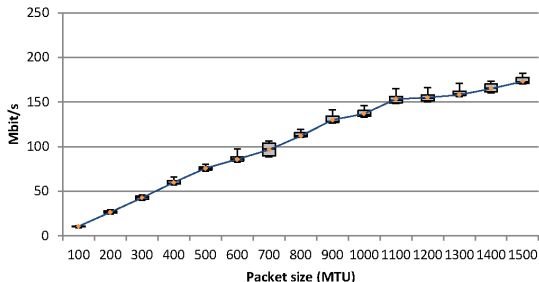
# Chart Types: Scatter Plot, Combinations

## ● ScatterPlot

- Usually used to display variables under control of experimenter (measurements)
- Efficient tool for showing correlations (positive, negative, null), usually combined with best fit line

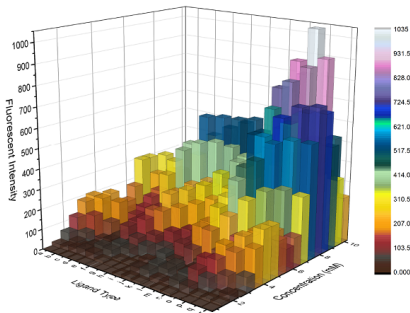
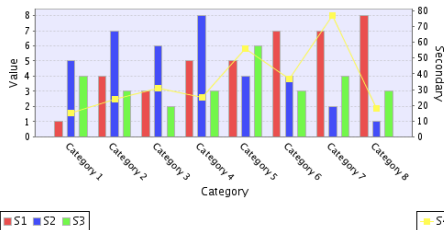


## ● Plot Combinations



# Special Charts

- Double-axis charts
- Logarithmic scale
- *Shiny 3D charts: usually not suitable for paper*
- Cheating with charts (3D effect, range/domain shift, ...)



- Show large amount of data in small space
- Have a good reason for having a table (need to show exact values)
- Data in the table must be easy to read and interpret
  - Font size comparable to that of the text (may be 1 pt smaller)
  - Minimize the number of lines: mainly use to separate headers (not each line)
  - Some experts say the vertical lines should be avoided completely
- The content must be self-explaining (with caption or table legend)
  - Any symbols used in the table should be defined (ideally in the caption)
  - Units in physical entities must be shown: e.g.  $v$  [ $\text{ms}^{-1}$ ]
  - Amount of text inside the table should be minimized
- In some situation, a table-dedicated page can appear
  - Table can be in landscape orientation (requires turning the page when reading)

# Bibliography and Citations

- Use  $\text{BIB}\text{T}\text{E}\text{X}$  if possible (i.e., every time)
  - Serves as a database of papers (possibility to combine with another tool as Mendeley)
  - Several styles available, easy to share, combine, archive (plain text)
  - Correct typesetting of references according to norms
- Think of issuing  $\text{BIB}\text{T}\text{E}\text{X}$  correctly if you change anything in the .bib file (multiple executions + latexing again)
- Editing the .bbl file generated by  $\text{BIB}\text{T}\text{E}\text{X}$  should be avoided
- The conference/journal template/guidelines usually provide/prescribe the bibliography style:
  - Paper description style at the end of the paper
  - Citation label style in the text ([1], (1), (Turing 1936) etc.
  - Order of citations (alphabetical, w.r.t. the year, appearance in the text...); depends on the community
- <https://kuk.muni.cz/vyuka/materialy/>

- Be careful when creating your B<sub>I</sub>B<sub>T</sub>E<sub>X</sub> database
  - Correct category: article, inproceedings, proceedings, inbook, book, manual, misc, unpublished
  - Correct items (given for each category): authors, title, year, pages, publisher, volume, issue
  - Correct spelling of names: diacritics (using T<sub>E</sub>X commands), names separated with “and”
- Abbreviations of journal titles
  - Good place to search for some more space
  - Official list: <http://www.ams.org/msnhtml/serials.pdf>
- Using “et al.” in citations (*Chicago Manual of Style*, Cambridge University Press)
  - Never for less than three authors
  - Three authors: when repeated (the first occurrence: all names)
  - Always for more than four authors

# Acknowledgment, Short Biographies

- Acknowledgment is usually required by many funding projects
  - Sometimes, the phrase used to acknowledge the institution/project is given
  - Be careful about the number of acknowledged project/institution number/code
- Individual persons or teams can be acknowledged as well
  - Not co-authors, but participated somehow (for example consultation, access to resources, data)
  - Acknowledged persons should be informed in advance
- Short CV is often required in good journals
  - Two paragraphs summarizing your education, affiliation, major achievement
  - You can easily find a “template” from a short bio of an English-speaking author (not much variance)
  - Small photo is sometimes needed



- Images, videos or similar material which is:
  - Submitted in order to provide additional information to the reviewers
  - Provided by conference/journal for download for readers once the paper is published
- The material is *supplementary not essential*
  - It might never be accessed by the reviewer/reader
  - Should only explicitly illustrate the results clearly and sufficiently presented in the paper
  - Should not contain anything that is not presented in the paper
- Videos are very often uploaded as a supplementary material
  - Recorded by camera or screen capture
  - Equipped with subtitles linking it to the paper (Kdenlive, iMovie etc.)
  - Usually limited file-size (rather small format, but keep sufficient resolution and quality)
  - Be very careful about codecs
  - The preparation takes usually quite a lot of time: plan ahead (test tools when you are not under deadline)

# Numbers, Units, Dashes, Abbreviations

- Every time use SI units if possible (Hz, m, s, Pa, N,  $\Omega$ )
- Avoid abusing math-mode for numbers and alphabet units
  - *Wrong:* GHz,  $ms^{-1}$
  - *Correct:* GHz,  $ms^{-1}$
- Discussions about the space between the number and unit: \, \;
  - `siunitx` package: `\SI{4.5}{GHz}` gives 4.5 GHz
  - *Wrong:* processor running at 4.3GHz; mean velocity of 5  $ms^{-1}$
  - *Correct:* processor running at 4.3 GHz; mean velocity of 5  $ms^{-1}$
- Make distinction between hyphen and dash:
  - - (hyphen): compound words: *science-fiction*
  - – (2× -, en-dash): ranges: *3–5 days*
  - — (3× -, em-dash): digression in a sentence: "—":  
*In this case—and only in this case—we propose to...*
- Frequent abbreviations (use `newcommand`, `def` in the header)
  - i. e.: “id est”, “that is”: `\def\ie{i.\,e.}`
  - e. g.: “exempli gratia”, “for example”: `\def\eg{e.\,g.}`
  - w. r. t.: “with respect to”: `\def\wrt{w.\,r.\,t.}`
    - *Do not repeat “to” after w. r. t.!*

# Using Symbols

- Define only those symbols you really need and use later
- Use `newcommand` or `def` to define symbols in the header:
  - Extremely easy to change a symbol for any quantity
  - $\LaTeX$  source becomes more readable (equations)
- Be consistent: e.g. **A** for matrices, **v** for vectors, *x* for scalars etc.
- Be careful with Greek letters:  $\rho$  vs.  $\varrho$ ,  $\epsilon$  vs.  $\varepsilon$  (`var-` prefix)
- Use `boldsymbol` to be sure you have bold Greek symbol (if needed):  $\lambda, \alpha, \epsilon$  vs.  $\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\epsilon}$
- Avoid double indexing (index of an index):  $A_{x_\alpha}$
- If dealing with many matrix, vector operations (tensors in general), consider *Einstein notation* (usually used in physics)
  - Automatic summation over indices repeated in the same term
  - $v_i = \sum_j A_{ij}x_j$  becomes  $v_i = A_{ij}x_j$

## Will be the subject of the last session

- Choose any scientific paper you like
- Focus on the way it is written
- What is good, what is not
  
- Upload into IS before the session
  
- The presentation is informal
- It should last up to 10 minutes