Purpose of Scientific Writing (and Why to Practice it)

The aim of **science** is to generate new and reliable **knowledge**. If one only **applies science**, perhaps without doing own research, one should gather and apply state-of-the-art knowledge rather than outdated knowledge.

Due to the aim of science, good writing is not a luxury, but an essential skill. It is essential not only for the dissemination of knowledge, but also for creating knowledge: During the writing process one is forced to be precise in all details. It is only during the writing process that gaps in the reasoning or lack of clarity become apparent. According to the philosopher Artur Schopenhauer, unclear writing is an indication of unclear thinking, and everything that can be thought can also be said in clear and unambiguous words. A similar quote:

> *"Whereof one cannot speak, thereof one must be silent."*
> (Ludwig Wittgenstein)

In a more general perspective, scientific writing is a process of **organizing and shaping information**. Compare writing to programming: When we write computer programs we apply the utmost care: The program must be syntactiacllay correct and must compile and run, furthermore it must be logically correct, and we spend further efforts on optimizing, structuring, etc. Shouldn't we apply at least the same care when preparing information for other *people* rather than for machines?

Scientific Argumentation

Some principles of writing follow right away from the purpose of science:

- Any scientific article must build upon the state-of-the-art, therefore authors are obliged to actively search for all relevant results related to the subject, to put their new findings in the context of existing results, and to apply the best known methods.
- All we can say in science can (and must) be said **clearly**. Any vague, evasive or ambiguous statements are not scientific.
- Be clear about any **assumptions** made. Also distinguish and make it obvious: What are proved facts, and what are only hypotheses, conjectures, etc.?

- Doubts are normal. Do not hide them. Frankly point out and discuss doubts.
- Almost all scientific results build upon other results which build upon other results ... It is impossible in practice to verify everything from scratch. Therefore other scientists need to **trust** published results. Also, society wants to trust science and be sure that published scientific claims are true, or at least reflect an honest and unbiased attempt to describe the truth.
- **"Mind the gap!"** Check whether all arguments are logically coherent: Every step must be **conclusive** and really follow from the given assumptions. Obviously, in mathematical proofs these are absolutely strict demands – otherwise mathematical reasoning breaks down. But also informal reasoning in any field of science should follow such standards as closely as possible.
- Substantiate every claim, and if you can't, remove the claim.
- Stick to all promises. Do not announce big things and then never come back to them in the text.
- It is also part of scientific thinking to **motivate** approaches and choices of methods: If alternatives exist, why has this one been chosen, and not another one that seems applicable, too? What are the advantages?
- Negative results are results, too! It can be important to convey that a certain approach does not work, in order to prevent others from repeating the same efforts in vain. (As a CS example, undecidability and NP-hardness results are valuable, as they show that certain algorithms, or certain efficient algorithms, cannot exist.)
- But do not describe events – a scientific text is not a diary. Fruitless attempts, detours, and falsified hypotheses should be mentioned only if they provide negative *results* and important insights about the subject itself, and not only an account of incidental, subjective mistakes.
- All scientific claims can be challenged, and authors must be prepared to defend them. Even excellent work invites criticism. This is nothing bad, it is an essential element of science. Try to anticipate possible questions and doubts of opponents, find the potentially weak points

and work on them. (Of course, at some point one must stop and release the text.)

- Scientific questions never end, therefore give an outlook: Say what could not be addressed, and what seem to be the most promising further questions.

An example regarding clarity: A popular phrase is "X is something like Y". This is not only too informal language, it also lacks a precise meaning. Instead characterize the relationship: "X equals Y but is described in a different way" or "X is an example of Y" or "X is a special case of Y" or "X is analogous to Y", etc.

Write for the Readers (not only for yourself)!

Many recommendations follow already from this principle. But we will elaborate on them in detail.

The first question to ask is: What is the intended audience, and what will the readers probably do with your text? Here are increasing levels of use of a scientific text:

- The reader wants to become aware of the subject and decide: Is this is an interesting or relevant matter for me? Do I want to read more?
- The reader wants to learn about the main results, methods, achievements, etc., at least on a high level.
- The reader wants to actually use what the article provides (e.g., methods), but without being forced to go through all details first.
- The reader wants to study the subject in all detail, perhaps for own continued research and development, and get fully convinced of the truth of all claims.

Also notice that there are different types of scientific texts, including:

- **original articles** reporting new results,
- **reviews** that summarize and discuss other work in order to guide potentially interested readers,
- **survey articles, monographs, and textbooks** that treat an entire field comprehensively, sometimes from a new or more general point of view.

Structure of Articles

Elements of an Article

In order to serve all aforementioned purposes in one document, articles usually have an abstract, an introduction, and a technical part.

The **abstract** is a short summary of the article. An abstract must be a self-contained document, comprehensible without the body of the paper. In particular, it must not contain references to any items in the paper. For instance, if one wants to cite some work already in the abstract, one cannot write "see [8]", referring to paper 8 in the bibliography. Instead one must give the bibliographic details of the cited paper, or if this is too long, give at least the authors' names and the publication year.

The **introduction** should give succinct statements about

- the subject of the article: what is it about?
- the scientific challenges: what precisely are the problems?
- the motivation: why is this important?
- the background and context: what related work has been done before?
- the main achievements: what precisely are the take-home results?
- their significance: why is this a big step?

After reading only the introduction, a reader should already have a clear idea what to expect from the article and feel well informed even before diving into the technical details. Also non-experts in the specific domain should be able to understand the introduction. Therefore, do not assume much prior knowledge.

A **conclusions** section is often put at the end. It should not merely repeat the earlier summaries, but wrap up the article, assuming that the reader is now familiar with the details, and give an outlook.

Structuring at all Levels

An article must be **readable** in all parts. Nobody wants to get stuck all the time and ask oneself questions like: "What is this now?? Did I miss something?"

This requirement may look obvious. But what does readability precisely mean? And then, by which means can we accomplish readability? Let us begin with some key criteria:

- The article is comprehensible.
- It has a good flow of reading.
- It is easy to retrieve specific information later on.

What supports comprehensibility?

- First and foremost: Clearly formulate in each and every sentence what you actually want to convey. Take the reader's perspective: **A reader can only see your text and nothing else.** A reader cannot "see" any additional thoughts you might have had in mind.
- Do not omit facts or logical steps that look trivial to you but are crucial for the reader to follow. The reader is, most probably, not that familiar with the subject as you are, while you are writing about it. Simply said: Be helpful!
- The text should tell a story. Avoid loose ends, distractions, and other passages that have no purpose, and so called "red herrings" that mislead the reader.

- **Introduce all terminology!** It is better to give too many definitions rather than too few: Already experts in a neighbored domain are in general unaware of many special terms and synbols.

- Use established terminology wherever possible. Do not invent your private terminology. But if, for some reason, you absolutely must, then explicitly define in what sense you use the words and phrases. You know it, but how shall the reader know, if you don't tell it? Also be aware that many words used as technical terms also appear in everyday language, but with a different meaning. It must be easy to recognize whether the usual word or the technical term is meant.

- Make sure that each definition is complete and unambiguous and really characterizes the thing, without leaving room for interpretation.

- Define each notation before it is used for the first time, not afterwards. Moreover, definitions must build on each other. Ask yourself: Here I define some concept Y that presumes another concept X – did I already explain X before? If you like, you may use a dependency graph as a tool to avoid explanatory gaps: Draw a directed graph depicting the dependencies between the most important concepts, and check your text against it.

- Give all entities names, in particular, introduce enough symbols for mathematical entities and refer to them. It is much easier to read and retrieve "x" rather than "the quantity mentioned in the previous paragraph".

- Conversely, introduce/explain all symbols that appear in mathematical formulas. It is pointless to "throw" a formula without saying what it is, or what the symbols mean. It does not harm to clarify even standard notation, in order to resolve possible doubts. (Example: Of course, there is no need to say that ∂ stands for a partial derivative and $\sum$ for a sum, but perhaps it is not always obvious whether "e" means Euler's number or some variable.)

- Once your notation and terminology is defined, stick to it, and use it thoroughly. Do not jump between different notations and synonyms, as this can easily confuse the reader. ("Is this Q the same as the thing named P before? But why is it now called Q instead of P? ...")

What supports the flow of reading?

- Do not make your sentences unnecessarily long and tangled, such that one must read them three times forth and back to grasp the meaning. Split such sentences. Ideally use a mix of short and reasonably long sentences.

- Let your sentences present the information in a logical order: subject first; important information first; or old information first, followed by new information.

- When pronouns “it”, “this”, etc., are used, it must be obvious what subject “it” or “this” refers to. The subject should have been mentioned just before. If there is too much text between a subject and a referring pronoun, the connection is lost (for the reader).

- Parentheses and footnotes are perceived as interruptions. Use them sparingly.

- Paragraphs are not arbitrary groupings of sentences. A paragraph should deal with one specific theme within the text. Ends of paragraphs should be the natural points where one can interrupt reading and resume later. Avoid “monolithic” paragraphs: very long and unstructured paragraphs are tiresome to read.

- Always make it obvious what you want to do next, especially in the beginning of a section. Do not jump to other themes in an unmotivated way.

- Always write connecting text: in the beginning of a section, around formulas, etc.

- If applicable: It may be good to explicitly announce that a passage can be skipped at first reading without losing the thread, for instance, passages with very special technical details.

- In mathematical texts, avoid line breaks in formulas.

- Read your final text, in order to check whether it has a good rhythm, otherwise edit further.

What supports information retrieval in an article?

- Write only sentences that are essential and carry information, remove long winded and commonplace statements and void phrases. Always come to the point.
- It can be helpful to write important terms in a different font, e.g., italized, when they appear for the first time. This signalizes: It is here where the term is introduced and explained.
- In mathematical texts, of course, definitions, theorems, etc., must be highlighted as such. Simple definitions may also be given within the text, but more complex concepts, as well as simple but central concepts, deserve a formal Definition. Decide in each case what is more appropriate.
- Ideally use the same numbering across all highlighted items. If you number them separately (for instance: Lemma 1, Theorem 1, Lemma 2, Lemma 3, Theorem 2, Corollary 1, ...), it is harder to find the items later by their numbers, as they lack a global ordering.
- Use informative and meaningful headings and subheadings. They should reflect the actual content of the section.
- Every figure, table, etc., should have a reference in the text, such that its role in the article is clear. But the actual content should be specified in a caption, not in the referencing text, such that one can immediately make sense of the figure or table. Be specific in the descriptions, e.g., say what quantity a coordinate axis represents.

Intellectual Honesty

- Respect intellectual property. Never pretend that others' work is your own work.
- **Plagiarism** in any form is a serious offense. It includes: verbatim copying of material without mentioning the source (also paraphrasing, i.e., copying plus modifications, without mentioning the source), stealing ideas, stealing program code, and so on.
- Acknowledge earlier work you build upon, not only in the bibliography. Also, when you discuss details in the text, acknowledge facts like "This

concept/method was introduced by ... in ..." Specify what your own work is, and what is inherited.

- Reported data must be true – this goes without saying. However, besides fabrication of data there are less obvious cases of improper data handling, such as omission of outliers and other unwanted data, and small "corrections" towards a desired conclusion. All preprocessing and postprocessing steps of data must be documented.
- Whenever feasible, data should be made public, unless this is prohibited by an employer or violates ethical principles like protection of privacy.
- Extensive data that do not fit in an article can be provided as some form of supplementary material. It should always be possible for other researchers to **reproduce** the reported results.
- When an image produced by others is inserted, make sure that you have the right to use the image in that way, and cite the source (author, etc.).

Finalizing a Text

During the writing process an author is, naturally, occupied by the contents and focused on the single text pieces. Therefore a final overhaul of the text in its entirety is always advisable as the last step before a text is released.

Finalizing a text is not a dull exercise but can substantially improve the quality. Ideally the final revision can even trigger new ideas. ("Now that I read it again ... well, I have never thought before about this particular point ...")

- During the writing process it is common to insert new pieces of text somewhere in the middle, to move or delete pieces, to change formulations, and so on. All these changes may destroy the alraady established structures in the text, e.g., cause interruptions and logical gaps, change the order of explanations, destroy cross-references. Inserted passages may lack their necessary context. Therefore it is important to read the final text again and check: Does the text still "work" as a whole? Is the structure coherent? Are the parts well balanced?
- Write all headlines of sections in the same style. Either capitalize them or not, but do not mix both variants without any system.

- Use a spell checker; it is annoying for readers to encounter typos that could be easily avoided.

- Often it is required to cut down a text to a prescribed length. It is amazing how much space one can save by rephrasing, different wording, simplifying formulations, removing redundancy, and so on. This can even make a text more readable.

- Is the tone appropriate? The text must be objective and not pretentious. Also avoid both slang and vagueness.

- A proverb in gastronomy says: "You eat with your eyes first." Similarly, the layout of texts is not unimportant. A bad appearance easily gives the impression of lack of care. ("If the author is so sloppy already with formal details, why should I believe that the scientific work is sound?") Please check: Does the final layout look good? Place tables and figures appropriately. Avoid layout elements that hurt the eye, for instance, a headline close to the bottom of a page, an almost empty page with only a few lines, or other ugly line breaks and page breaks.