DAT315. Scientific Writing

Purpose

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science**, perhaps without doing own research, one should gather and apply state-of-the-art knowledge rather than outdated knowledge.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 this purpose of science, good writing is not a luxury, but it is 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 the details. It is only during the writing process that gaps in the reasoning or lack of clarity become apparent. *Unclear writing is often an indication of unclear thinking.*

Scientific Argumentation

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

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.
- Clearly distinguish: what are proved facts, and what are hypotheses, conjectures, etc.?.
- Doubts are normal, but 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.

- Be clear about any assumptions made. 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.

- 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?

- Also negative results are results! 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.

- 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 insights about the subject itself.

- All scientific claims can be challenged, and authors must be prepared to defend them. Even excellent work invites criticism; this is nothing bad, but 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 should not occur in a scientific text. Instead characterize the relationship precisely: "X equals Y but is differently described" or "X is an example of Y" or "X is a special case of Y" or "X is analogous to Y".

Use of Personal Pronouns

This is more a stylistic remark. Authors of scientific texts usually say "we" rather than "I", even if the text is written by a single author. This "we" can be understood as an "impersonal" pronoun that supports an objective tone. While the authors are behind their articles as real persons and responsible for the content, authors are not "part of" the actual scientific content. Next, it is uncommon to address the readers directly: simply avoid the pronoun "you". (Note that *this* text is not a scientific article, so we will say "you".)

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:

- Becoming aware of the subject and decide: Is this is an interesting or relevant matter for me? Do I want to read more?
- Learning about the main results, methods, achievements, etc., at least on a high level.
- Actually using what the article provides (e.g., methods), but without being forced to go through all details first.
- Studying the subject in all detail, perhaps for own continued research and development; getting 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 you want to cite some work already in the abstract, you cannot write "see [8]", referring to paper 8 in your bibliography. Instead you have to 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, now 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 annoyed by reading, getting stuck all the time and asking 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!
- 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?

- Make sure that each definition is complete and unambiguous and really characterizes the thing, without leaving room for interpretation.
- Give all entities names, in particular, introduce enough symbols for mathematical entities. It is easier to read "x" rather than "the quantity mentioned in the previous paragraph".
- Define each notation before it is used for the first time, not afterwards.
- Once your notation and terminology is defined, stick to it, and use it thoroughly. Do not jump between different notations and synonyms.
- The text should tell a story. Avoid loose ends, distractions, and other passages that have no role.

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.
- 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 very long and unstructured paragraphs; they are tiresome.
- 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 announce that a passage can be skipped at first reading without losing the thread, for instance, passages with 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, in order to signalize: 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.

Bibliography and Citations

Virtually every scientific work builds on earlier work or is at least related to earlier work. An author almost never starts from zero, and often the author "sits on the shoulders of giants". The main purpose of the bibliography is to put the article in context. Possible reasons to cite a paper in the bibliography are:

- The article builds directly upon the cited work, uses their results, and extends or improves them.
- The article is not a direct continuation of the cited work, but the cited work has introduced the concepts or developed key results.
- The cited work is about different subjects that are however closely related. It is mentioned as delimitation or as general background.
- The cited work is a standard reference, for instance, a common textbook or survey.

In the first two cases citation is mandatory. Other authors can even feel offended if they come across an article where their work is not cited although it clearly should. However, one may miss references for various reasons: the publication venue was not well known or not accessible, the title did not hint to the subject, etc. In fact, even a thorough keyword search in literature databases does not guarantee that all relevant work is found. Often the same subjects have been studied independently under different names. Therefore omissions are excusable, but one should at least make a serious effort to find all relevant and up-to-date work.

We list some further rules:

- Papers are cited in the text either by their number in the bibliography only, or by the author name(s) and the number in the bibliography.
- Obviously, a paper cited in the text must also appear in the bibliography. But also the converse is important: A paper should be in the bibliography only if it is cited in the text, because otherwise it is not clear how that paper relates to the present one.
- Only truly scientific work can be cited, but not, for instance, Wikipedia articles. While they can be very good introductions and serve their purpose, they do not meet the standards of scientific publications.

- Many articles are available online, but it is not common to mention links to the electronic versions in the bibliography. Give only the bibliographic data. Only when a cited article is not officially published, one should give the web links. But then also mention the date of access, because these documents might change.

- Different citation styles are in use (just look into different articles to see examples), and they have their advantages and disadvantages. The only hard demand is that all citations and bibliography entries should be written in a uniform style. Bibliography entries are usually copied from somewhere, and then it is easy to forget to adjust the styles.

- Text pieces of other authors can be cited literally, however: One must clearly mark quotations as such, and mention the source. Failing to do so is **plagiarism**.

- Really look up what you cite, do not simply copy from other bibliographies without thinking.

Readers may use the bibliography to get more background information: look up details of earlier results, look up concepts, definitions, and basic facts that are well known to experts but not to a novice reader. This also leads to an advice for **reading** a scientific paper: If you are not familiar with things that are apparently assumed to be known, then a good approach is to trace the main references.

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 ...")

- Is the structure still coherent and the parts of the article well balanced?

- 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.
- Is the tone appropriate? The text must be objective and not pretentious. Also avoid both slang and vagueness.
- 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 strange line breaks and page breaks. A bad appearance gives the impression of lack of care.

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 with 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 it 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.