Writing Theses and Other Scientific Documents

Clemens Simmer, May 2017

1 Introduction

Advance in science is only possible if we communicate our findings to others. Usually this communication is done in writing; a written document maximizes the potential impact on science because there is no limit on the number of recipients. What is often overlooked however – and probably every scientist has experienced this - science itself evolves while writing. Writing down our ideas forces us to structure our thinking into units like sentences and paragraphs, which we – and anybody else – can check for logic and truth.

In the following I try to summarize my experiences in and ideas on scientific writing resulting from my own writing and even more from reviewing scientific documents ranging from bachelor, master and PhD thesis, proposals to funding organizations and articles submitted for publications to scientific journals. I will start with the overall structure of such documents (Section 2), dwell quite a bit on the internal structure and logics (Section 3), discuss how figures and tables are best included and referenced (Section 4), and how external information sources used in texts should be referenced (Section 5). Section 6 summarizes my ideas and experiences on writing English text as a non-native English speaker; some ideas formulated here are, however, also useful for writing German text I believe.

2 Structure

Scientific documents like papers in a scientific journal, theses for PhD, master or bachelor studies (and to some degree also applications to funding organizations for research projects) usually follow a similar structure:

Abstract/summary
1. Introduction
2. State of the art (optional, can be part of introduction or theory)
3. Theory
4. Model description (optional)
5. Data (optional)
6. Model development/experiment description/workplan (optional, can be part of the introduction)
7. Results
8. Discussion
9. Outlook

References
Appendix (optional)

The **abstract/summary** should precede any scientific text – and it is the most important section, because it is meant to start your communication with the reader. This is what the reader will read first, and on the content of which she/he will decide to read more – or not. The abstract must be short (less than a page) and summarize the stated problem, the tools/data used to solve it, and the overall results. This section – which usually does not have a section number - must not contain any references, figures, tables, and also - if possible – no acronyms.

The rest of the structure is of course guided by the overall purpose of scientific texts, namely

- to formulate the problem at hand or a hypothesis including a motivation why it is important to solve the problem or to decide about the truth of the hypothesis (**Introduction**, or if more detail is necessary an additional chapter on **state of the art** may be considered**)**

    - The **introduction** section should get the reader interested. She/he should also get an idea, where she/he can find relevant published literature concerning the problem, and what ideas these authors had about the problem. And the reader should get convinced, why more research must be done, what knowledge about that problem is still missing, and what and how you contribute with your work. You should at the end of the **introduction** section roughly describe, how you approached the problem, i.e. the methodology. In project proposals, the latter is usually cast into a separate – the most important - chapter called "**workplan**". In theses and journal articles, you might consider – depending on the required detail for explaining your executed workplan to add sort of an "**experiment description**" after the theory, model and data sections. At the end of the **introduction** section you should summarize, what the reader will find in the rest of text, i.e. how the text is structured into sections and what they contain.

- to introduce the reader to the necessary theoretical background of the research topic, so she/he can understand on which theories your research is actually based upon (or you plan to base your work upon in a project proposals) (**theory, state of the art** sections)

    - This chapter is not always needed (and it also can be part of the **introduction** section) – you have to decide yourself. If you want to derive a new theory or model, or extend an existing theory or model, you should introduce the reader to this theory or to the physics of the processes, you want to model. If you derive a new measurement method for a variable, here you have to explain e.g. how the variable to be estimated, creates the information in your measurements.
    - You should restrict theory to the material really needed in your work or to understand your results. You should not list and explain an equation, which you do not need later and which is not necessary to derive another needed equation.

- to explain the existing tools and/or data needed or used to do your research (existing **models**, **data sets** used in your research)

- to explain additional tools and/or data needed which required (your) considerable extra work (**model development**, **experiment**, **observations**,…),

    - This chapter is only necessary, when you need considerable additional own scientific work before you can address the core scientific problems stated in your introduction. It is already a "result" chapter, contains your original work and should be described with all necessary detail.

- to show and explain the results from executing the work plan stated in the introduction (**results**),

    - This is the main part, the core of your text, where you detail the results of your central work. This can be several chapters, if necessary or appropriate.

- to discuss the implications of the results in relation to the stated problem and published literature or hypothesis (**discussion**),

    - Here you come back to the introduction section (and – if it exists – the state of the art section); you discuss your results in view of what was known before you did your work. You compare your results with findings of other authors, other ideas and possibly other hypotheses. This is a very important chapter, because mostly here you concentrate on your scientific reasoning – you show, that you are indeed a scientist (*Wissenschaftler*), meaning you have created new knowledge.

- to summarize the results and state their implications for the future (**conclusions**, **outlook**).

    - These chapters will be read by most of your readers in order to judge, whether it is worthwhile to read your text at all. They want to know, whether they will learn something new and important for them (**conclusions**). And they want to know, what you would like to do, if you have more time or unlimited funds (**outlook**)…

The (not-numbered) **reference** section contains all the references to other material mentioned in the text. Follow one style and provide enough information, so everybody can find this text. Try to avoid references to sources, which are not reviewed (grey literature, internet). For internet sources you have to provide the date, when you acquired the source.

The **appendix** is usually not needed (as the name indicates), and should contain material not needed to follow the main text. You can summarize here e.g. common statistical methods, figures to which you refer to in the main text in summary (if you discuss a figure in particular in the main text, the figure must also appear in the main text), computer code for documentation, or a lengthy derivation of a formula.

This is one (ideal) structure of a scientific text, but you should not deviate too much from it (or have very good reasons to do so). More often than not, however, not yet experienced authors have difficulties, to follow such a structure because of many reasons (awareness of those might prevent these difficulties):

- The problem and/or the hypothesis have not been (or sometimes could not be) formulated adequately and concisely enough in the beginning – so people often present more an itinerary of the work done. The reader gets bored because she/he does not know what the author is heading for and thus, why to read it at all.

- The scientific problem and/or the hypothesis planned to be addressed at the beginning of the work has actually not been solved by the work which has actually been done – so people tend to write excuses for not having reached the stated goals (or blame others or certain circumstances). So the reader quickly feels, that she/he will not learn anything new. Why should a reader be interested and read on?

- The original scientific problem and/or the hypothesis has changed due to other or so far unknown developments (this happens very often – that's the common way of science!) – so people tend to describe in length what all they have done and how the change occurred.

- …

As a result, the thesis or paper structure often tends to mirror more the development of the work - but not the interesting and important outcome – the new science. Thus, authors should better try to follow the structure stated in the beginning – they should reformulate problems, hypothesis and realign the work plan according to the work actually done and the results finally obtained, because

- the reader is mostly not interested to learn all the usual difficulties and bad luck connected to scientific work (most time the reader is a scientist and knows about this anyhow too well), and

- you want to convince your reader about the value of your results ("sell" your results - and yourself).

Thus, the reader is mostly NOT interested in how your project really evolved over time. She/he usually also does not want to know exactly, how many wrong paths you took until you found your way (or your problem) and achieved a useful result – she/he wants to know and fully understand your new results and their implications. Of course “wrong” paths might also contain scientific value – in that and only that case they should be presented and discussed. And you can always add an extra chapter (e.g. a sub-chapter in the *conclusion* section) on the actual development of the work, explain initially stated wrong hypotheses or original goals not reached in the end, and how it all evolved.

Sections/chapters can be split up into subsections/subchapters, but do not exaggerate! Any subsection/subchapter should contain more than one paragraph. Try to stick with a maximum of three levels; a subsubsection 1.1.1 makes only sense if there is also a 1.1.2.

3. Internal structuring and logics

Besides the general structure, there are other things you should consider in order not to make the reader angry and let him/her stop reading. If you really want (and you should do want!), that somebody else reads and understands your text, you have to structure and formulate it in such a way, that the reader is able

- to understand the content of each individual sentence,
- to guess from the chapter headings what they contain (but do not make them longer than one line)
- to follow a “roter Faden” in every paragraph and chapter and from paragraph to paragraph and from chapter to chapter,
- to start almost everywhere in the text and understand what you want to say, and
- to keep his/her interest in your text.

It is almost impossible to explicitly tell, how to write a scientific text by many more rules. It is much easier to write down things often done wrongly. In the following I summarize the “Dos” and ”Don’t Dos” of scientific writing from my own quite extensive experience in reading other people’s stuff as sort of a check list.

1. Check each sentence individually, if it is logical and if it actually contains information. If not just delete it. Especially if you use words like “because”, “thus”, “when”, etc. check if the conclusion a reader might draw when reading this sentence is – not only grammatically - correct. These words assume a logical connection between sentences or parts of sentences, so be sure that the stated logic really exists.

2. Check logics also on the grammatical side. If you use “this”, “it”, or “therefore” etc. check whether the reader will not be drawn to the wrong conclusion, because you are NOT referring to the last sense-making sub-sentence. Best is to completely avoid starting sentences with “this” or “it”. Better formulate what you really mean.

3. Check logics between subsequent sentences within a paragraph. All sentences of one paragraph must be logically connected (constitute a *Sinneinheit* in German). Otherwise split the paragraph.

4. Avoid tautologies. A tautology is a sentence, which repeats in its sub-sentence the content of the main sentence just with different wording. Obvious tautologies are e.g. “The method number one compares better to the measurements and thus leads to more realistic results.” Tautologies can, however, be much more subtle. I have read texts, which needed more than twice the necessary length only due to tautologies. This makes readers tired (and angry).

5. Paragraphs should always separate logical units of texts. If you start a new paragraph with "From this follows…" (the best is to avoid this completely!) you do refer to the complete content of the whole preceding paragraph – not to the last sentence. Be sure that this is what you want. If you want to refer only to the last sentence of the last paragraph, you should consider not starting a new paragraph.

6. Do never repeat literally (e.g. numbers, the color or style of lines, the relative positions of lines, etc.) the content of figures or tables or their captions in the main text. Your text must INTEPRETE figures and tables! The figures and table are just means to prove or corrugate your conclusions stated in the main text. Try to avoid starting a sentence with "Figure nn shows…".

7. Be as short and concise as possible. I once read a sentence cited from a book about the style of writing, which is even more valid for scientific writing:

"… a sentence should contain no unnecessary word,
a paragraph no unnecessary sentences,
for the same reason that a drawing should have no unnecessary lines,
and a machine no unnecessary parts …".

In summary - you should use the fewest words possible to get the idea across.

4 Figures and tables

Figures are usually added in order to visualize observational data, model results and the results of their (statistical) analyses – in the text you have to explain, what they should take home from the figures. The figures should have exhaustive explanations in their figure captions below the figure, i.e. the reader should be able to understand what is shown on the figure without consulting the main text. Thus, never repeat figure captions or larger parts of it in the main text – the main text should only interpret the figure or summarize its information content. Be sure the figures and their axis texts are large enough (be aware that in journal articles figures might become quite a bit smaller than in your submitted text). Use colors ONLY when necessary. Use the same style for figure axis descriptions, numbers and legends in the whole document.

Most tables structure MANY numbers – i.e. it makes no sense to construct a table for four numbers. Like in figures – do not repeat its content in the main text, and put sufficient information in the table caption (usually above the table!), so the reader understands the table without reading the main text.

EACH figure and EACH table must be referenced and discussed in the main text.

5 References

Especially in the **introduction**, **state of art**, **discussion**, and **conclusions** chapters you will have to refer to the work of others, usually documented in reviewed papers in scientific journals. Decide for one style (copy it from one of the papers you read) and use it consistently throughout the text. I personally prefer to name the authors in the text (e.g. Meier and Müller, 1999) instead of just numbers (e.g. [6]), because the reader might know these persons or even the paper already and thus does not need to go to your reference list (which distracts from reading your text).

- Do not cite any literature, the content of which you do not discuss or summarize at least with one sentence.
- Do not cite papers, which you have not read yourself at least partially. In can be quite embarrassing, if you cite a paper because others have cited it and used a wrong reference (wrong year, title, page numbers or even content!). This happens more often you might think!
- Avoid the use of internet pages as references for scientific statements, because they are volatile and are not peer reviewed. If you do add the date when you consulted the page.
- The papers actually read by you and used in your thesis as reference material reflect (at least in part) how deeply you have worked yourself into the problems addressed. It is not the number of papers cited, which counts but how you have included their content in your text. However, for a bachelor/master/PhD thesis I expect a minimum of >10/30/50 of such references, excluding manuals, textbooks, and internet sources.
- If you use/copy text from other sources literally, you have to put this text in quotation marks and cite the source. Be aware, that nowadays software is available and used, which compares your writing with any other existing text.
- Even if you just reformulate the content of text from another source you still have to cite the source.

6 “don’t dos” for Germans writing English text

Being not a native English speaker or writer, I can only provide feelings about what is right and what is wrong which has accumulated by reading many English texts. Everybody has its own style, and my German style can of course be sensed immediately by native English speakers also in my writing. Nevertheless there are some constructions, which – when mixed with non-English thinking – might lead to misunderstandings or just wrong conclusions by the reader.

1. Better never ever start a sentence (or even a paragraph) with “however”. I know that there are also native English people doing so frequently, and then it might be ok. But for sure your text is better readable and automatically much more logical, when you avoid it. The problem with “however” is, that it refers to a statement (or a summary of statements?) made before. By using “however” at the beginning of a sentence, these statements are often difficult to identify. When used within a sentence, however, the meaning is usually much more clear.
2. Try to strictly obey the sequence rule ‘Subjekt-Prädikat-Object’ (SPO, hard especially for Germans).
3. Never create a sentence with more than one subordinate clause (Nebensatz, also a very good idea for German texts!).
4. Try to use active formulations instead of passive ones, whenever possible (also a good idea for German texts). This way texts becomes shorter and easier to read and understand.
5. Avoid qualifying words if there is no real reference; if something is “good” or “better” you have also to tell the reference compared to which “good” and “better” is to be understood (also a good idea for German texts). Better quantify you judgements.

7 Text editing and word processors

If you are a perfect writer and you need nobody to correct your text, and if only you are writing the text alone i.e. you have no coauthors and no supervisor, whom you want to get involved in helping you to do the writing – you might consider skipping a word processor and go ahead directly with a type setting tool like e.g. LaTeX. In all other cases you should use a word processing system, which allows you and everybody else involved to make changes in the text, which you can accept or decline (track changes option). In doing so reduces the communication work (and time) at least by half (also the work of your coauthors/supervisors) and avoids misunderstandings. I am not a friend of Microsoft WORD because of the commercial aspects, but I use it almost exclusively because almost all of my collaborators, with whom I write papers, use it. LibreOffice is an alternative but mixing WORD and LibreOffice more sooner than later will end up in problems. So, first discuss with your co-authors and/or supervisors what system they prefer to use.

When writing scientific papers be aware that most journals will typeset your text with their own type setting tool; thus your text will "look" different from the document you sent in when it is finally printed. And they employ people much better trained than you in this respect to do the type setting for you. Some journals even request that you submit your text as a WORD document.

A word of caution: word processing schemes, which are also doing the type setting, like Microsoft WORD or LibreOffice (YSWYG – you see what you get) are luring first writers to concentrate from the beginning too much on type (and figure, tables, …) setting instead of on the scientific content of the text. In a way this also true for LaTeX; and you can spend days on "beautifying" your document. Be aware, that in the end it is your text – and only your text – which is considered when your work is evaluated for scientific content and value. Usually only the text will contain the scientific innovation.

8 Conclusion

Good luck!