

Pleasing your readers

A brief guide to creating appealing, clear and easy-to-read scientific documents

Jan Lagerwall, jan.lagerwall@lcsoftmatter.com
September 2022

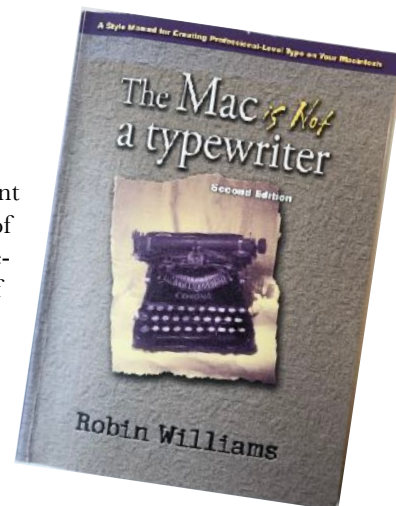
1. Introduction

A skill that rarely gets the training it deserves in academic education is that of writing texts and laying out, or designing, a complex document containing texts, pictures, equations, tables etc. A few decades ago, few people needed to know the principles of layout, because important documents were always typeset by professionals. Today the whole typesetting process can be done in any word processor, so these jobs no longer exist. Unfortunately, the evolution of word processors was not paralleled by an evolution of education in layout. Instead, the general attitude seems to be that layout does not really matter. It is very rare to see somebody teaching it within a university education (at least the scientific ones). The result is as easy to see as it is sad. Poor layout significantly reduces clarity and often introduces risks of misunderstandings, sometimes quite serious.

This is an unfortunate situation, in particular for you (I am assuming that the reader of this document is a student of higher university education), because whether you like or not, it is more than likely that you will be required to write many documents throughout your professional career, regardless of whether this will take place in academia, in industry or elsewhere. The rewards for your efforts will *not* just depend on the content, but in fact very much on how you present it. Failing to give your document a professional layout and/or being careless about language and spelling will inevitably give your readers a bad impression. As a certain producer of deodorants put it in a slogan: *Because first impressions last.* In other words, even before s/he gets to the actual content of your report/article/dissertation, the reader (one of the first of whom is likely to be a referee) will have a negative attitude to your work. In cognitive science this is referred to as anchoring bias and it is something that we cannot get around. Our very first impression sets our expectations for the rest.

This guide attempts to help you ensure that the reader's first impression is good, such that s/he has a positive attitude to your document. (You are then profiting from what is called 'the halo effect'.) Be-

yond what I have learnt through many years of writing, reading and reviewing, I base much of what I write here on the lovely little book "The Mac is not a typewriter" by Robin Williams (not the comedian). It's a bit old by now but the design recommendations are just as valid today as they were when the book was written.



2. Choosing and learning to use your tools

In order to ensure that a clear and appealing document will be the result of your writing efforts, you have to start by making a few choices and asking yourself a few questions. First of all, *choose* your word processing software! Don't just use what everyone else is using or what was installed by default on the computer you're using; make an active choice. Try out a few different word processors; there are many more available than the famous ones, for Windows, Mac as well as Linux, and many are quite low cost or even free. Ask yourself if you like what you see:

- Is it intuitive to work with?
- Is it aesthetically appealing, logically constructed and easy to find help and documentation?
- Does the document look the way you want it, on the screen as well as on paper?
- How does the program handle importing of images of various kinds (including .pdf) and which are the export options?
- Do the equations look good? Does it take ages of clicking your way through a screen menu to write an equation or are there keyboard shortcuts so you can get even complex equations typeset rapidly?
- How does the software handle large documents like a Ph.D. thesis? Can you separate a large document into several files (for instance chapters) which the software will put together for



you in a correct way in the end, also generating a table of contents and perhaps an index?

- How does it handle references and/or interact with external reference databases?
- Are there good solutions for collaborating with others on a document using the software?
- Will it automatically update the numbering of your figures and tables if you add a new one or remove one in the middle of the text?

To help you in building your opinion, check out the user reviews on web fora to see what errors, problems, clumsy solutions as well as good things are pointed out. Try repeating the problems discussed and see if you agree with the critique. Don't hesitate to write to the developer if you have questions; customer support is an area where small developers often outperform the big companies by huge margins.

You will obviously need to invest some time to do this kind of testing but—believe me—it is an excellent investment. Remember that you will be writing many texts over the following years. Whether this will be a pleasure or a nightmare largely depends on if you made a good choice to begin with. Some writing software (like the various incarnations of LaTeX) may have a high learning threshold, but once you have learnt it, writing may be much smoother and more efficient than with a software with a lower threshold.

EVEN A BORING AND UGLY DESIGN IS STILL A DESIGN.

As a general advice, do not make the mistakes of thinking that (1) you “are not designing documents” (you are, whether you like it or not, whether consciously or unconsciously, since the document that you produce *will have* a certain design and you are imposing this onto your readers) and (2) that the software will take care of writing style for you. The one exception to the latter point is LaTeX (with which it is fairly complex to personalize the style, however) but most software will give you limited help concerning style. Even if you use templates, you need to learn a few things about document design to use them properly.

In fact, without your active intervention, a software like Word will generally give you an unappealing document that can be difficult to read, and it may even introduce errors that you did not make. Auto correction may turn every letter following a period upper case, so if you punctuate an abbreviation (which you should) that does not ap-

pear at the end of a sentence, the software will cut your sentence in two. Or try writing a letter-based list like (a), (b) and (c) without getting the last item header replaced by a Copyright sign. The text will be very confusing if you do not correct these software-created mistakes. The situation becomes critical if you work in a language different from that for which the auto correction was optimized.


Regardless of which software you choose to work with, you must spend some time learning it. The time invested in the testing phase is actually not wasted, as it automatically also contributes to learning. For instance, you need to learn how to define paragraph and character styles and make sure that the right one is applied at the right point. I am often surprised at how often I receive Word documents where the whole document is in the paragraph style “Normal” and no character styles are defined. Instead an impressive number of manual modifications appear throughout the document. Not only must it have been very painful to format this document, but it becomes particularly vulnerable to formatting errors introduced because of a platform change, for instance between Windows and Mac, between old and new versions of Word, or between different language settings. Collaborating on a document without defined styles is doomed to give problems. And if at some point you wish to make a general layout change throughout a long document, it's going to take you a while if you did not define your styles. If you are not aware of the styles functions in your software and how powerful they are, you are working very inefficiently and missing out on a great aid.

Another thing you must learn in your software is how to handle figures. I will say a few words about how to treat the actual figures in section 10, but here my concern is the importing of figure files of formats like .jpg, .png, .svg, .tif or .pdf and then giving them the right size and position in the document, as well as a caption that sticks to the figure. Moreover, make sure that you learn how to get the surrounding text to float around the figure, because good layout requires you to be able to choose the position of the figure regardless of changes in the text.

Finally, do learn the keyboard shortcuts for the most commonly used functions. It saves you a lot of time and allows for much better ergonomics than when you have to use the mouse all the time. Your arms and neck will thank you.

3. Paragraphs and sentences

Now that you have chosen your tool (and learnt to use it to a sufficient degree), it's time to look at the



formatting of your document. An issue many inexperienced writers have trouble with is paragraphs: how long should a paragraph be and what is a good place to break paragraphs? A *very* rough guideline could be to aim for 5 ± 2 sentences per paragraph. However, the main guiding principle must always be the logic of the text; you can never replace critical thinking with a number. If you change topic, it's probably a good idea to start a new paragraph. If you have a very long discussion about a topic, you still need to break that down into multiple paragraphs. So look for natural "breathing pauses" in your text, where you can start a new paragraph to help your reader.

Once you have mastered the separation of your text into paragraphs you will encounter the next issue: how should you layout your text such that the beginnings and ends of paragraphs are visible but not disturbing? The common default setting of most word processors is to have neither indent (the small space at the beginning of new paragraphs that I use in this document, unless they follow a heading) nor extra space above a new paragraph, making it very difficult to spot where a paragraph ends and the next one begins. In fact, if a paragraph happens to end with a line where the text extends all the way to the right, it is impossible.

The classic way to indicate the beginning of a new paragraph (open any professionally printed English novel you may have at home and you will see this layout being applied) is to use an indent, as I do in this document. I personally like this the most as it gives the text a certain graphical liveliness rather than just being a continuous flow of words. Remember, however, that the indent should be used functionally (it is a general rule that layout should be functional): if a paragraph is preceded by a heading there should be no indent because it is not needed. The heading indicates that a new paragraph will follow. By defining paragraph styles in your word processor, including the sequence of styles (for instance, a heading style is followed by a style without indent, which is followed by one with indent, which is followed by the same style) the presence or absence of indentation is easily automated.

My taste is of course not shared by everyone and many people prefer the completely straight left paragraph of a non-indented text. This is perfectly fine as long as a new paragraph is preceded with some extra space. The lazy way to do this is to simply hit an extra Return (or Enter) between two paragraphs, but much more elegant and efficient is to program the paragraph style to add something like six points of space (half the height

of a regular line of text) above the paragraph. This also has the great advantage that it is very easy to adjust the amount of space later on, should you need to fit the full text in a certain number of pages. Just like you should not use indent after headings, you should not combine an indent with extra line space; one way of indicating the new paragraph is enough.

The width of an indent should, according to classic design guidelines, equal that of the capital letter M. Incidentally, this is also the width of the long dash, consequently called M (or em) dash. When using this dash—for instance to insert a comment within a sentence—remember that there should be no space before or after it. The slightly shorter dash, –, is called an N dash (or en dash) because—you guessed it—it has the width of the letter N. It is used, e.g., in stating ranges, such as a distance range (4 – 7 m) or a time span (20 – 30 s or July – September). When used in this context, the N dash is separated from the surrounding text or numbers by spaces, and it is pronounced “to”. The N dash is also used for writing compound adjectives, like in “the Santa Fe–Chicago flight” or in “high risk–high gain projects”. In these cases no space surrounds the dash. Both N and M dashes are distinct from the hyphen, -, which is used to distribute long words over multiple lines (hyphenation) or as a minus sign.

What about the text near the right margin? Should it be straight (justified text), as in this document, or should it end at varying position depending on the length of the words? Justified text can look very elegant, but if you choose this layout, be sure to hyphenate your words as required to avoid long white spaces between words or weirdly separated characters. This will otherwise inevitably happen when a long word comes towards the end of the line, the likelihood of which increases proportionally, the greater number of columns you are using.

A paragraph consists of sentences, and also the sentence length is very important for readability. A common phenomenon is that the first version of a text contains sentences that are much too long, because you are formulating as you are thinking. This is always the case in my own writing and I am by now so aware of this habit that I always scan a fresh text of mine for sentences that would benefit from cutting in two. Most often, the clarity increases and the text gets less heavy to read if long sentences are broken up into several smaller.

One final thing needs to be stressed here, namely that a sentence *always* begins with a capitalized word. This may sound trivial, but it is not un-



common to see inexperienced writers of technical texts write sentences that begin with a variable, which can be lower case, Greek, etc. If you do this, it is impossible for the reader to be certain that this is the whole sentence; was the beginning of the sentence perhaps cut out by mistake? Thus, never start a sentence like “ l increases linearly with $t \dots$ ” or “ α is independent of $x \dots$ ”. One way of avoiding the problem is to use the meaning of the variable for starting the sentence, even if the reader already knows it. Thus you could write: “The length l increases linearly with $t \dots$ ” and “The angle α is independent of $x \dots$ ”.

4. Table headers, figure captions, and numbering

Another common problem is table headers and figure captions. The general rule is that a table header is placed *above* the table (that is why it is called a header) whereas a figure caption follows *below* a figure. I am not certain for the reason behind this convention, but I imagine it is because we read from the top to the bottom. When a reader gets to a figure, it is good if their eyes immediately see the figure, then read the caption. The caption gives additional explanations but it is not required for a first look. In contrast, a table without explanation of what the data show is often rather meaningless, and therefore the reader should start by reading the table header, looking at the content of the table afterwards.

Remember that, in a scientific document, each figure and each table should be referred to at least once in the text; if the figure/table is so unimportant that you do not naturally mention it in your writing, then it has no reason to be included. In this document I *do* have a few figures that are not referred to and they also do not have captions or figure numbers. This is because their function is mainly decorative. They are there to break up a long streak of text with some graphics. This is fine in texts of this kind, but in a scientific article such figures would generally not be acceptable. I also don't refer to the single numbered figure in this text, but this is because the caption is there for demonstration purposes.

Tab. 1. An example of a table summarizing data from an experiment. Note the proper formatting of variables and their units.

Sample	T (°C)	U (kV)	t (s)
A	30	12.4	6.2
B	40	14.3	5.4
C	50	16.6	5.0



Fig. 1. Appropriate sentence length depends on the medium. Cartoons are not the medium of long and complex sentences.


Tables and figures should both be consecutively numbered, such that all references to them are unambiguous. The numbering and renumbering of these elements can become a nightmare in a large document if your word processing software does not include a good way to handle dynamic cross references. Word can do the job with limited success, while a program like Apple's Pages (the software used for producing this document) in its current incarnation fails entirely. In Pages you have to type every figure and table number manually, just like the references to them. You can imagine the fun if you realize towards the end that you need an additional figure at the very beginning!

5. Equations, numbers, variables and units

Equations are considered part of a paragraph, even if they are often typed on a separate line. For instance, if I need to make a definition:

$$S = \sin x \quad (1)$$

I continue the paragraph without indent after the equation, since the equation is part of the paragraph. Note that an equation given a separate line (with some extra space above and below the equation) is generally numbered, often with the number in parentheses near (but not necessarily at) the right margin. This way you can easily refer to the equation later on in the text. If you write a small equation that will not need to be referred to again, for instance $F = k \cos^2 x$, then you typically do not give it an extra line, and no number. The ease in writing equations, getting them positioned appropriately in the text, numbering and referring to them, differs greatly between word processors. Apple's Pages recently got much better as it now incorporates a LaTeX equation editor, but the numbering is still not automated. LaTeX is generally best in class for handling equations overall.



Note in the above equations, as well as in the header of Tab. 1, that variables (S , x , F , T , U , t) as well as parameters (k) are written in *italic* typeface. In contrast, units ($^{\circ}\text{C}$ for degree Celsius, V for Volt, s for seconds) and prefixes (k for kilo, M for mega, G for giga, m for milli, ...) are *never* written with italic typeface. This universally accepted convention is actually important, as the wrong typeface can change the meaning entirely. For instance, if you write 3 km, the reader interprets this as the distance three kilometers, whereas 3 *km* will probably mainly confuse the reader. It could be interpreted as a relation stating that something is three times the mass of an object, weighted by a factor k . And would 2 *GB* mean a file size of two gigabytes, or twice the gravitational constant multiplied by the magnetic field strength? Mathematical functions, like sin, log or ln are always written non-italic, separating them nicely from the variables they operate on, thus $\sin x$, not *sin x* or *sin x*.

It is also worthwhile paying attention to the fact that upper and lower case letters often refer to different things. Note in Tab. 1 that two column headers have a variable represented by the letter T. It is not a coincidence that the first column, where data have the unit $^{\circ}\text{C}$, uses upper case T , because the convention is to use upper case T for temperature. Lower case t is instead used for time, hence the unit is seconds (s). Sadly, it has become a rather frequent error to see K used for the prefix kilo in many texts (often even without including the unit for which it is a prefix!). This is truly problematic, since prefixes appear where there are units, and the convention is that K is short for the temperature unit Kelvin. Thus, KM is *not* kilometer but Kelvin mega, whatever this is supposed to mean.

In the context of numerical data, it is important to mention the issue of the number of significant digits. I have often been surprised at how little attention this issue gets, sometimes even from experienced writers. If I write $m = 2.598,328$ kg, I mean that I am certain down to the single-milligram scale about the mass. The object really has a mass of 2.598,328 kg and not 2.598,327 kg or 2.598,329 kg. I cannot give this many numbers of precision if my certainty is only on the gram scale! Many people just read off a reading from an instrument or software, forgetting to think critically about where the certainty in the value ends.

The polymer chemistry/engineering community has a peculiar bad habit belonging to this category, and this is to indicate molar mass in g/mol, without using powers of 10 or suitable prefixes, regard-

less of whether this makes sense or not. (Adding to the confusion, it is not uncommon in this community to drop the unit entirely, and treat molar mass as if it were dimensionless.) Thus, a certain sample of polyvinyl-pyrrolidone is frequently said to have a molar mass of 1,300,000 g/mol, which is problematic if you think about the enormous dispersity in molar mass of synthetic polymers. For a physicist (like me) it is obvious that the proper way of indicating the molar mass would be 1.3 Mg/mol, but this makes just about every polymer chemist or engineer I encountered cringe, without giving a better reason than “nobody writes like that”. I have never understood the argument “because everyone else does a mistake I must also do it” but I note that it is not uncommon. I find it counterproductive and it certainly does not comply with the ambition for critical thinking that a university education should strive for. If you absolutely cannot use prefixes when indicating very large or very small numbers, then please at least use powers of 10. Thus, a good compromise is to write that the molar mass of the sample is $1.3 \cdot 10^6$ g/mol.

And since we now arrived in the land of super- and subscripts, I have to write something about the increasingly common bad habit of ignoring these formatting tools. Water is *not* abbreviated H₂O, the short-hand for helium-3 is *not* ³He, a sodium ion is *not* symbolized with Na⁺, a parabolic function is *not* written $y=x^2$, and a good notation for the starting time of an experiment is t_0 , *not* $t0$ (or $t0$ or T0). It is incomprehensible to me how so many students and researchers totally ignore proper super- and subscripts in texts and (in particular) in presentation slides. If you really want to look ignorant or lazy, this is an excellent way of giving this impression. If you do not *know* how to get super- and subscripts with your word processor, then you clearly did not do your homework regarding section 2. *Please*, do your readers a big favor and make sure you get super- and subscripts right.

There is a second important source of confusion that we have run into in this section, and that is the use of period (.) and comma (,) in indicating numbers. In English writing, the period is used as a decimal stop, thus $\pi = 3.14$. In many other languages, the comma instead has this function, thus in Swedish, German, French and Italian, the corresponding equation would read $\pi = 3,14$. The proper choice is to use the convention of the language that you are writing in; thus, if you are writ-



ing in English, please write 3.14, aber auf deutsch sollte es 3,14 sein.

To add to this confusion, the *other* symbol is often used for a different function, namely to separate orders of magnitude. Both English and French do this, but with inverted symbols. Thus, the speed of light in English is $c = 299,792,458$ km/s mais en français, elle est $c = 299.792,458$ km/s. Pretty confusing, right? Well, c'est la vie in our polylingual and polycultural world. ;-) It's actually very simple once you know the conventions, but you need to learn them once. If you go back to the discussion of masses above, you may notice that all the numbers I gave comply with the English way of writing numbers, since the language of this text is (primarily) English.

To wrap up the discussion on numerics, it is worth mentioning that good style recommends writing out numbers lower than 20 (10 in some traditions) in letters, unless they are followed by an abbreviated unit. You would thus write that a process took six seconds or 6 s, but not 6 seconds.

6. Headings

A common layout issue in written documents as well as in presentation slides is that authors do not care about the look of their multi-line headings. It is preferable to fit a heading on a single line but sometimes it is not possible. In the latter case, one should *choose* where the line is broken, because it both looks bad and worsens readability if just a single word or two end(s) up on the second line (these are called orphaned words). Look at the title and decide where a line break is appropriate from a logical and aesthetic point of view, and then make the line break by typing a forced line break at that point (on the Mac it is typically Shift + Return or Alt + Return, depending on the software).

To demonstrate the point, look at this heading written unconsciously and consciously, respectively.

Long headings should be broken manually at the right point

Long headings should be broken manually at the right point

Which looks better and is easier to read in your opinion? The effect is even more striking in a presentation slide, where the title often appears with very large font. The same thinking about symmetry can be applied in the vertical direction at the end of a multi-column text. You will note that professionally laid out texts with multiple columns typically have all columns of a final page that does not fill the full page ending slightly higher than the

regular page bottom. This way they are balanced, rather than having one very long column and then the adjacent being just a few lines.

For a heading we also need to choose the way it is given emphasis. Common choices are **bold-faced** type, *italic* type, underlined text, larger font size, ALL CAPITAL LETTERS, or perhaps a different font. Sometimes several of these choices are combined. Emphasizing by underlining or by using all capital letters should generally be avoided. Underlining belonged to the typewriter age, when the type could not be changed, leaving underlining as the only way to emphasize text. Why this has lived on decades after the death of typewriters is a mystery. If you capitalize letters you make the text much more difficult to read, since you lose the vertical variation in extension of the letters. When we read we do not really process text letter by letter, but our brains do advanced pattern recognition, scanning blocks of text much more rapidly than we could possibly parse text one letter at a time. This pattern recognition becomes more difficult if you remove the vertical size variation, explaining why any text in all capital letters longer than a single word or two gets tough to read.

7. Choosing and spelling your words correctly

When it comes to choosing words, there are (at least) two basic requirements for a text to be enjoyable and serving its purpose. On the one hand, the terminology must be adequate, clear and descriptive. On the other hand, there should be sufficient variety. These requirements are not in conflict with each other; for anything but special technical terms there are synonyms. If you use one and the same word for an element of your discussion all the time, your reader will find it rather boring and possibly even annoying. Don't be afraid of using a thesaurus (at least on the Mac, an excellent thesaurus & dictionary opens if you right-click a word and choose "Look up"). Even if you master the language you are writing in perfectly, some external help can be extremely useful to find inspiration or get out of a deadlock when it comes to choosing words.

Note, however, that synonyms of technical terms should be avoided for clarity. For example, if you define a period as *pitch* at one point, but then refer to it as *period* later on, a reader who is not an expert in what you are discussing may think that you are now considering a different concept. Or if your text deals with nanorods and suddenly you mention cylinders; are you talking about the nanorods or some other cylindrical shape? For all

technical terms, decide which term you will use and then stick to it throughout your document.

Do use your word processor's spelling and grammar checkers! If you give a document to a referee (whether an unknown journal referee, your thesis advisor or someone else) and they find spelling or grammar errors that the word processor would have detected, you will have given that person the impression that you are very lazy. Remember the anchoring cognitive bias and think about what the consequences will be for how the referee judges the rest of the document. Also make sure to be consistent about *which* English spelling you use. Set the spell checker to either English or American spelling (or another version if you prefer), and then make sure that this type of English is used throughout. If you are writing for a specific journal or book, you may not have a choice about which English spelling to use; check the guidelines from the publisher.

It is so easy to run the software spell checker that it should *always* be done during or directly after writing. In general, never give a document to another reader before you have gone through it carefully yourself, assisted by all tools that your software offers. You want your proofreader to help you improve the text from the best of your ability, not fix errors that you were too lazy to fix. Don't waste the time of people you depend on: you will make them disappointed and they will have less time to help you with the important issues.

Here I would like to bring up a very common spelling error in German and Swedish texts (the problem surely exists in other languages as well), which is to some extent 'inspired' by the English language. This is the bad habit of separating single words into components. While in English a vacuum cleaner is two words, in Swedish and in German it is one word (*dammsugare* and *Staubsauger*, respectively). If you separate the components (*dammsugare* or *Staub Sauger*) you give your sentence a different meaning. Actually, it will usually *lose* its meaning in a strict sense, but in some cases it can still be a correct sentence, only conveying quite a different message than you had in mind, sometimes with quite embarrassing consequences. The serious problem with this typing error is that no machine spell checker will discover it, so you have to be careful about this issue yourself.

8. Some words on grammar

I will not write much about grammar here, as I assume that you will refer to proper books about the English language if you do not master its grammar well. There are, however, three issues

that are so common that I feel that they need to be mentioned in a text like this. One concerns the incorrect choice of indefinite article ("a" or "an"), another the incorrect use of apostrophes, and the third the singular/plural forms of imported words.

Perhaps the problem with the wrong choice of indefinite article comes from English teachers who simply teach *rules* rather than *reasons*. One such rule may be that you always use "a" in front of a consonant and "an" in front of a vowel. These types of rules, that say nothing about the reason for the rule, are not very helpful and they can be quite misleading, as in the case of the indefinite article usage. The *reason* is simply convenience: it is much easier to pronounce a vowel sound following a consonant sound and vice versa. Therefore a noun, adverb or adjective beginning with a vowel *sound* is preceded by the indefinite article "an", whereas a noun, adverb or adjective beginning with a consonant *sound* is preceded by "a".

Thus, although 'h' is a consonant, we write "an honest person", not "a honest person", because the 'h' in honest is not pronounced. Likewise, although 'e' is a vowel, we speak of "a euphemism", not "an euphemism", because the word euphemism is pronounced starting with a consonant sound. To get this right, you obviously need to know how the English word is pronounced. A fairly common problem is that non-native English speakers know the proper rule, but they get the pronunciation wrong. Thus, I have sometimes seen "an history" rather than "a history" in writing from people with a native language where the 'h' is rarely, if ever, pronounced.

Note that this also applies to abbreviations that you expect the reader to read as such. Thus, if you introduce the abbreviation LT for Language Trainer, and you then write "a LT", you are expecting the reader to replace 'LT' with 'language trainer' when reading. This is very unlikely to happen, and probably you were yourself thinking 'ell-tee', not 'language trainer' when writing it. Therefore, you should write "an LT", not "a LT" since the *sound* of the abbreviation starts with a vowel. For the sake of choosing the indefinite article, it is in fact irrelevant whether the *written* word/abbreviation starts with a consonant or vowel; what matters is how the sound starts.

The issue with apostrophes is illustrated well by the text below, that you can find on t-shirts, stickers and on many social media threads. The word

Irony is when someone writes "Your an idiot".



“your” is a possessive pronoun that signifies that something belongs to you. It is surprisingly frequent (not least among native English speakers!) to mix this up with “you’re” which might be *pronounced* similarly but has a very different meaning. When you use the apostrophe, “you’re” is a contraction of “you are”. A similar problem is the mix-up of “they’re” and “their”, the former a contraction of “they are”, the latter the possessive pronoun for plural third person. Correct would be “They’re upset because their car was stolen”, but not “Their upset because they’re car was stolen”.

Equally common are difficulties in getting the apostrophe related to the genitive case right. While you do use apostrophe-s for indicating belonging to a name or a noun (this is, for instance, Jan’s text and you may be reading it from your computer’s screen), this is *not* correct when you wish to indicate belonging to the pronoun “it”. Thus, if you have been speaking about an animal’s appetite when describing a life science experiment, and later wish to replace “the animal” with “it”, you should write “its appetite”, not “it’s appetite”. The latter would actually mean that you are defining something as “appetite”, since “it’s” is a contraction of “it is”. Note also that in situations where you normally would use apostrophe-s for indicating genitive, but the word or name itself ends with ‘s’, you use only the apostrophe. Thus, we may speak of Chris’ car, not Chris’s car. The same holds for the plural genitive case, where we use just the apostrophe, without the ‘s’. Thus, we could write about the “the mice’ appetite”.

There is a fairly simple and safe way to avoid these problems, and that is to just avoid contractions in your writing. For scientific texts, contractions like “it’s”, “they’re” or “he’s” are generally not appropriate style, as such contractions are considered too colloquial in most contexts. So if you simply write out “it is”, “they are” and “he is” and reserve the apostrophe for the correct genitive uses, you have largely avoided the problem. On the other hand, often it is wise to avoid the genitive case in scientific writing, and write “the appetite of the mice” rather than “the mice’ appetite”. It can yield clearer and more precise language and it again sounds more professional and less colloquial.

Regarding singular and plural forms of imported words, the issue applies strongly to scientific English, which is rife with words imported from, e.g., Latin and Greek. It is quite common that the plural forms are used even when the word is in singular, which is generally incorrect. For example, a single item of the plural words *phenomena*, *spectra*,

media or *data*, is called, respectively, a *spectrum*, a *phenomenon*, a *medium* and a *data point* or (rarely used) *datum*. The use of *data* in singular is now so common that it, at least in non-scientific texts, is often accepted, treated as a mass noun (like *information*). However, “a phenomena”, “a spectra” or “a media” are all wrong, without exception.

Let me end this section by pointing out that it is surprisingly frequent to run into inconsistencies in number in many texts. If you choose to discuss items in plural, then please make sure that all your words are in plural, and vice versa for singular. If you write “the samples is placed in the chamber” or “these experiment is repeated” your readers will cringe, but even the more subtle “such materials are an example” is troublesome. For the latter case, a way out is to write “such materials constitute an example”.

9. Punctuation

A document of this kind cannot avoid the topic of punctuation, which is a constant source of confusion and frustration for many. (Now I am not referring to the numeric use of period and comma, which I already dealt with above.) When should you write a comma and when should you not? What is the function of a semicolon? Can I connect these two sentences through a colon or should the first be ended by a period? These are questions many inexperienced English writers, native and non-native, struggle with, and getting it wrong can give rise to awkward reading. I cannot go through the rules in deep detail in this document, but I’ll give a few basic guidelines.

The period in American English, full stop in British English, indicates the end of a sentence (or of an abbreviation) and it is quite rare that writers place a period where there should be none. More common is that they *omit* a period where there should be one, simply by allowing a sentence to grow too long. Sometimes a period is incorrectly replaced by a comma, colon or semicolon. For any of these punctuation marks, there has to be a strong relation between the text before and after the comma, colon or semicolon. For instance, a colon is appropriate when you follow up a piece of text with a text that does not flow smoothly from the first, but fills an explanatory, qualifying or completing function, as in “Never forget: the reader will be misled by your mistakes.”. The standard explanation for when to use a semicolon is when you cannot decide between period and comma. This is not a very precise guideline, but it is actual-

ly reasonably fitting. If you do not feel confident with the use of the semicolon, just avoid it.

The most critical issue is probably the use of the comma, since it is used extensively, and it is needed. Yet, if placed in the wrong place, it can ruin the readability. A good baseline tip is the recommendation of my former Ph.D. candidate Larry Honaker (native English speaker from the USA): if you would naturally make a pause when speaking, then write a comma. I add some more specific examples of how commas are used, but this is far from an exhaustive list. First, you often use a *pair* of commas to insert an explanatory text within a sentence, as in “The girl, whose lunch I had stolen, looked very angry.” Note that you can remove the two commas and everything in between, and the sentence still works out. It’s just that you were left without some quite significant information. When using commas like this, remember that they *must* be used in pairs, before and after the inserted text.

Another common case of pairs of commas is around “for example” or “that is”, which are often abbreviated with “e.g.” (for Latin *exempli gratia*) and “i.e.” (for Latin *id est*), respectively. We could thus write, e.g., “I had stolen her lunch, i.e., I left her without anything to eat.” If you wonder why I had two periods towards the end of the last sentence, it’s because the first period belongs to the sentence that I was citing, within quotation marks, whereas the second period ends my own sentence.

A very common use of commas is to separate items in lists, as in: a scalpel, a knife and a dagger. You may have heard of “the Oxford comma”, which would have been inserted before the “and” in the above list. The occasion when the Oxford comma should always be used is where there is an ambiguity about the meaning if you do not have the comma. I recently ran across a good example on-line: “We invited the rhinoceri, Washington, and Lincoln”, versus “We invited the rhinoceri, Washington and Lincoln”. The former case (with Oxford comma) would mean that Lincoln and Washington as well as the animals were invited, whereas the second one (without Oxford comma) would mean that we invited the two animals, which are called Washington and Lincoln. Search for the texts and you’ll find a nice graphical illustration of the difference (I cannot include it here since I have not requested permission).

10. Some core guidelines for figure preparation

And speaking of permission for figures, it is really very important to be aware of the fact that all figures come with usage rights. While some pictures

on the internet can be used without reference to the source, this is not the case for most graphics. In some cases you are allowed to use the graphic if you credit the source, in other cases you must ask for permission. This of course holds for graphics from articles and textbooks as well. If you get permission, this generally comes with very strict guidelines on how you should refer to the source and the copyright holder. Be sure to follow these guidelines! If there is no mention of the copyright holder, it is an implicit statement that the document author holds the copyright. Thus, if you did not make a graphic yourself, you must write who made it. The cover of Robin Williams’ book on the first page of this document is actually a photo that I took of my own physical copy of the book; had I just copied an image of the cover from the internet, I would not have been allowed to include it. Sometimes quite subtle differences can have important consequences.

The focus of this document is on text, but since scientific documents almost always contain figures, I should say just a few words also about how to prepare figures, so you can avoid some critical mistakes that I have seen too often. First, do not forget that a photo, and most times also a drawing, has a fixed aspect ratio. This means that if you change the width, also the height must change by the same factor. Surprisingly, it seems that the default setting in many applications that are often used for making figures (some more words on which software to use for this below) do not have the aspect ratio locked as the default setting. I see many persons changing the width more than they change the height (or vice versa) using such software, not realizing the sad consequences for the image and the information it conveys.

If you distort the aspect ratio of a photo of a face of your friend, that person will certainly not be happy seeing her-/himself with a fattened or slimmed down face, and the person hopefully makes you aware about your mistake. But even a scientific image with distorted aspect ratio is highly problematic. Not only does the concept of a scale bar become useless, since you now have a different scale in the vertical and in the horizontal direction, but the shape of the objects in the image has changed. An ellipsoid is a significantly different object from a sphere, since it is no longer minimizing its surface, so if you show a spherical liquid droplet as an ellipsoid, because you have squished the photo of it vertically or horizontally, you are giving the impression that your liquid does not obey the laws of physics (or simply that some



strange conditions break the symmetry and allow for the non-spherical shape). If you do not think that this has consequences, just imagine playing tennis, golf or (European) football with a non-spherical ball. As for objects with discrete rotational symmetry, like a hexagon or pentagon, they will lose that symmetry when distorted, with angles that are no longer constant throughout the shape.

So, please, lock the aspect ratio of all objects that have any kind of physical significance, which includes all images like photos or micrographs. If your software does not allow to lock it by default, often you can get it temporarily locked by keeping a modifier key like Shift or Ctrl pressed while reshaping the object with your mouse. Check the manual or help file of your software.

And concerning software, I strongly encourage you to use proper graphics software for making your figures. While software like PowerPoint and Keynote can be used to draw graphics, this isn't what they were designed for. You simply don't have the control of the result that you will with a good dedicated graphics program. In any case, please do not draw figures with your text processing program, like Word or Pages. You will often get poor drawing quality, thanks to the limited tools for graphics. More importantly, you are seeding great problems when editing your document. If a figure needs to move to a new page, some component you drew may stay where it was. If you need to change the size of the figure, the text in it will not scale simultaneously. And even if you think this might be a convenient situation at the writing stage, if you are preparing an article for a scientific journal, the type setting staff will require a dedicated graphics file for every figure once your paper is accepted. So save yourself work at the end by preparing the figures properly from the start.

Concerning graphics formats, if you do not know the differences between vector (infinitely scalable) and bitmap (pixelated) formats, and between compressed and uncompressed graphics, then I recommend that you learn about this. Formats like .jpg, .png and .tif are bitmap graphics and .jpg is always compressed whereas .tif is often uncompressed. The .pdf format is fundamentally a vector format, but it can embed bitmap graphics, which are then, of course, no more scalable than when they were free-standing. Another very important aspect is if you can have a so-called alpha channel, which allows you to make the background of the graphic transparent. This is the case for .png and .pdf, but not for .jpg.

11. Sharing your document electronically

Finally I would like to bring up one computer-technological issue that many writers are not paying attention to. When you create a document

with many images, the file size can get very large. This is particularly the case if you include high-resolution photos and scale them down to small size without first reducing their resolution to values that are appropriate for their size. So, if you have a series of images taken with a state-of-the-art microscopy camera and you compile them into a table at full resolution, you may create a document with effective image resolution of several thousand dots per inch (DPI). A normal value for good printable resolution is 300 DPI.

The best way to avoid this problem is to make a copy of your images and save them at 300 DPI in the size that you are going to use them. This will render your word processing file *significantly* smaller in size and thus much less heavy to work with for your computer, which means you will enjoy much faster response when editing or moving around in your document. If you work with LaTeX, your compilation time will reduce greatly. The same thing holds if you are preparing a presentation in software like Powerpoint or Keynote.

In case you did not reduce the original image file sizes, then at least please reduce the final document file size before you share it with others. For instance, if you export your document into .pdf format and wish to e-mail to your advisor, to co-authors or to other colleagues, please check how large the file size is before you send it. If it is more than 10 MB, you should compress it first. There are many excellent pdf compressors on the market, for a variety of operating systems, and there are even websites that offer pdf compression for free. I very frequently use a wonderful Mac app called "Compress PDF" and it costs only \$4.99. It's a *very* worthwhile investment, because it often reduces a pdf file 10-fold or more, without any visible consequences for the images (at normal magnification).

12. Closing remarks

In the end, remember that you can always have the ambition to challenge the rules and come up with your own ideas. Indeed, this is the essence of human evolution. So am I saying that you just wasted your time reading through this text? No; while you may well have the ambition to go your own way right from the start, you should do so first when you master the rules. With time and a reasonable degree of devotion to the topic, you will certainly reach the stage where you can freely design according to your own mind. However, if you don't really care about designing your documents, or if you find the topic boring, you will do best by simply learning the rules (and the reasons) once and for all and then following them. Good luck with your writing!