WRITING

REPORTS

AND

THESES

INSTITUTE OF INFORMATION SCIENCES AND TECHNOLOGY
MASSEY UNIVERSITY
PALMERSTON NORTH
NEW ZEALAND

True Ease in Writing comes from Art, not Chance,
As those move easiest who have learn'd to dance.

Alexander Pope
Essay on Criticism

WRITING

REPORTS

AND

THESES

INSTITUTE OF INFORMATION SCIENCES AND TECHNOLOGY
MASSEY UNIVERSITY
PALMERSTON NORTH
NEW ZEALAND

Writing Reports and Theses

Many books propose sets of "definitive, all-embracing" rules for writing Honours reports, Masterate, and Doctoral theses. Unfortunately, their advice is usually heavily weighted towards the research techniques characteristic of the author's subject, and can mislead readers from other areas. Those authors who overcome this innate subject bias usually go too far in the opposite direction, and say everything and nothing. A few even manage to achieve both subject bias and complete vagueness.

To avoid these pitfalls, and in recognition of the restricted nature of my readership, I have deliberately oriented this document to the sorts of research project undertaken by Computer Scientists. However, it concerns *writing about*, rather than *carrying out*, Computer Science research, and thus it emphasises the construction of documents, not programs, Data Flow Diagrams, Structure Diagrams or other standard CS items. *Furthermore, it does not replace discussions between student and supervisor.*

By the time you start work on your report, you'll have successfully completed the practical part of a research project; you will be abreast of its subject, both at a theoretical level and at the practical level of the project. In short, you will be an expert in the area. Unfortunately, it isn't enough merely to possess expertise; you must also be able to communicate it clearly and accurately, in the wider context of the whole subject.

Your research report is, therefore, much more than just a summary of your work, written so that your supervisor - a very special reader, who knows the context of the research - can dream up an appropriate mark. In fact, it is an ambitious writing project that must be far more accurate, coherent, scholarly, and searching than anything most of us have ever produced as undergraduates. As technical or scientific specialists, some of us have achieved excellent examination grades in spite of only rudimentary language skills. A few of us even take considerable, though perverse, pride in this. At your current academic level, however, regurgitating previously absorbed information is no longer enough; your research generates new information which is available from no other source, and poor writing can only inhibit its recognition.

Specifically, poorly integrated, non-grammatical writing disturbs the smooth flow of information from writer to reader, even if it includes all the relevant facts. The writing should allow the reader effortlessly to perceive connections between related points, but it too often forces the reader to infer them from a welter of misrelations and *non sequitur*s. In short, poor writing obscures the picture.

These are questions that may prompt you to structure your writing carefully and well:

How do I decide whom to address?

How do I subdivide the material?

How do I translate my *gestalt* into sequential text?

How do I avoid common writing errors?

How do I refer to The Literature?

How do I decide whom to address?

The style of your report will depend largely upon your *target reader*, the reader whose knowledge and understanding you keep in mind when formulating your arguments, and whose needs and interests determine what is to be included in or omitted from your report. As there are many prospective readers, including directors of sponsoring firms, your supervisor, and later researchers, identifying this special one is not simple.

Let's start by observing that your report is primarily an academic document, recording your achievements for the benefit of academic posterity. Most academic progress results from sequences of small advances made by independent researchers. Only very rarely does a single great intellectual leap advance science by a hundred years. Your report, however insignificant, must stand as a building block in the growing edifice of knowledge. It's not a company report, a newspaper article, a marketing brochure, or a piece of popular science.

So we've narrowed it down: your target reader is *posterity*. Great. Now all we have to do is characterise posterity and address it. Doesn't sound much fun. Let's try anyway.

In the case of a research report, (by contrast with a paper in a journal, for example) posterity consists of independent researchers who may wish to repeat your work to confirm your conclusions, independent researchers who wish to base their project on your conclusions, later students following up your research under the same supervisor, and, perhaps surprisingly, yourself. It's often useful to be able to refer, in later publications, to conclusions presented in an earlier one, without having to repeat the possibly lengthy arguments that led to them. That is, you may later wish to cite your own report, in, for example, an article in a journal. This implies something about the minimum standard of argument in your report.

One might also surmise that one's supervisor would be a good reader to target. After all, she or he will be marking the report. However, your supervisor will generally already be thoroughly familiar with the general subject area and the details of your project, and the pair of you will be able to converse about them in a sort of private code, using informal diagrams, and referring to shared personal experiences. When you write a report, you need to distance yourself from this informal style of discussion.

Taking these somewhat contradictory factors into account, therefore, I suggest that the most appropriate readers to target are your successors in the project. These people are most likely to be students working for the same supervisor, and enrolled for the qualification which you are seeking; i.e. if yours is an Honours project, then an Honours student is likely to follow you; if a Ph.D. project, then a Ph.D. student. Pitch the report at a level at which a student embarking upon a follow-up project, with appropriate previous education, will be able to understand it. That is, don't write course notes for an introductory undergraduate paper (though, if they are central to your project, you might reiterate the definitions of terms that such a paper introduced). At the other extreme, don't write the report so that only someone who has already completed a similar project can understand it.

It may seem surprising, but pitching your discourse at this level is a worthwhile stratagem in your pursuit of a good class of Honours. After all, your examiners are interested in determining how well you understand the introductory and surrounding material, as much as (and in some instances, more than), reading a record of what you actually did. Writing a good general introduction to a topic is a good way of demonstrating your knowledge of it.

How do I subdivide my material?

Choosing the format of a large report or thesis is not an easy task. Some parts*have* to be present. Others are optional, depending on the subject area. Still others are optional, depending on the particular style of research you have undertaken. Before you start writing, and probably after you've started, you and your supervisor should discuss how various aspects of your work fit into these classifications.

There is general agreement, however, on the broad outline of a research report in a practical subject area. These components would normally be present:

Abstract: A succinct, and essential, summary (no more than about 300 words) of what you've accomplished in your project. Abstracts - often collected and published separately - enable other researchers to determine whether your work is relevant to theirs. Your abstract should therefore concentrate on your own research, and should not act as an introduction to the field nor present arguments that led to your conclusions.

Contents: A ready reference to the location of each Chapter, and Section within a chapter (see my comments about decimalised numbering systems, though), listing these in order of appearance, with associated page numbers. Word processors are good at generating Contents sections automatically. In some documents, it is worth including lists of Figures, Tables, or both, in a similar format.

Introduction: A "funnel", leading the reader from general, broad, considerations of the field, to the particular problems *your* research concentrated on. It normally contains:

a summary of the field, (suitable for an appropriately trained person just entering on the specialised type of research you have undertaken), including the *general* problem dealt with in your research.

an introduction to, and statement of, your *thesis* - the belief which has helped you formulate your approach to the problem. In revolutionary work, this statement of your "philosophy" will probably have to tread a fine line between being arrogant (undesirable) and being provocative (desirable). The thesis, or an abbreviation of it, might well replace the sterile word "Introduction" as the title of the first chapter.

a brief statement of the *particular* problem you worked on; your thesis would have suggested its character. Careful work on phrasing this description accurately will help you later in deciding whether a sentence, section or paragraph elsewhere in the report is central or tangential to the topic.

a plan of the structure of your project, describing the major components of the research and showing how its various threads are woven together. This does not mean an expanded CONTENTS section.

Specification of the problem: a complete discussion of the *particular* problem which is the subject of your project, with an indication of how it fits into the general subject area.

Previous work on this particular problem (literature review): comparison with original approaches developed in this project; reasons for your choice.

Report of system design and development: particular problems inherent in the approach; modifications to the approach, engendered by consideration of these problems.

Experimental evaluation of your original system or algorithm or technique: a report on experiments designed to test the development, with their results and the interpretation of those results.

Conclusions: Argument at a more general level. More than just a *summary* of what you've done, your thoughtful conclusions should reverse the funnelling action of the introduction by setting your development in the context of the research area as a whole, showing, for example, how it is an improvement on previous methods*. You may include suggestions for further work here, but don't fall into the trap of thinking that you can use this section to excuse you from completing the practical work of your own project. Markers will spot this from a long way off, with their eyes closed.

Appendices: More peripheral, supporting, material, usually extensive, which you do not refer to *directly* in the report. If you wish to refer directly to a table of data, a program listing, or a formal standard, etc., then include it (or, if it's too big, an extract or summary) in the body of the report, not in an appendix. You might even need to do both; summarise it for inclusion in the body *as well as* including it in full in an appendix. Just don't use the excuse that it is too big to go in the body of the report as a reason for not summarising it. You'll only be trying to avoid doing some necessary hard work, and the quality of your report will suffer.

Similarly, if you wish to refer specifically to parts of your system's code, include these parts as figures in the body of the text, where they stand some chance of being read. Markers are rarely interested in complete program listings, as- I estimate - 95% of most program source code is just mundane housekeeping. If later researchers may wish to use the programs, include a disk in an envelope attached inside the rear cover, or assemble a copy for the Department to archive and note this in the report. Otherwise, program listings only make a report heavier. Omit them.

Citations (in the body of the text) and References (at the end of the document): the Harvard style for the citation of references and the list of the references themselves is to be followed - see the final section of this document. *All* sources which have contributed to the writing must be acknowledged. It's generally considered that the probability of two people using identical word sequences to describe something becomes vanishingly small when the word sequence exceeds five words in length. So if you quote a word sequence that's more than six words long, you should include a

* or is not - this is just as valid a conclusion to reach, though hardly as satisfying.

citation. And citations aren't just used when we quote someone directly; if you paraphrase (reword) someone else's ideas, that work needs to be cited too.

Acknowledgements - whatever thanks you wish to express to whoever helped you towards your goal.

The divisions above could map directly onto the chapters of the report or thesis, but need not. For large projects, particularly Ph.D.s, system development is likely to be iterative and require a finer division of the chapters than I've suggested here. These would be particular to each project, and cannot be outlined in such a general document as this. Further, the above outline relates to the phases of a system development project, by contrast with, say, a literature search, which would omit discussion of your system's design, implementation and evaluation. It would probably have sections containing a critique of previous work or suggestions for future research topics.

Deciding whether to divide a particular section of the work into chapters is very project-specific, but might be done on the basis of length. For example, one wouldn't often expect a chapter to contain fewer than, say, five pages of text, or more than forty. You certainly shouldn't treat these numbers as absolute limits but it would be worth looking critically at your chapter divisions if you find you're writing chapters with lengths outside this range.

I'll discuss the report's total length in the next section.

How do I translate my *gestalt* * into sequential text?

When writing a large document, you're grappling with a number of problems simultaneously - everything from getting your spelling and grammar right, to final consolidation of your understanding of the theory, and decisions about how big it all ought to be. We'll leave the more technical, niggling, difficulties till the next section and deal here with the all-pervading problem which underlies all the other, more particular, points you may think are the real trouble. Unfortunately, it's so nebulous that many writers don't recognise it at all until it is explained. It relates to the structural difference between an author's preliminary mental model of a complex subject area and the final printed document describing that subject area.

When you've been working in an area continuously over a period, *all* its aspects come to the forefront of your mind, and their interrelationships are clear. By the time you complete your research, they form a mental, often multi-level, network which *you* can jump into at any point and skip around with practised ease, performing leaps of understanding which baffle ordinary mortals.

This type of integrated, overall mental model - to which I apply the psychological term *gestalt* - is fine for you when you are busy performing research, but, at write-up time, you face the task of mapping its two-, three-, (or even more-) dimensional structure onto the strict linear structure of a report, a sequence of chapters, sections within chapters, paragraphs within sections, and sentences within paragraphs. Accomplishing this mapping unobtrusively is the essence of your task.

Unfortunately, the task is frequently complicated by your model's inadequacy. It is easy, or at least common, for a researcher's mental model to contain a very high-level view of the problem area, and thorough familiarity with the practicalities of solving it, but nothing much in between. This is particularly true of Honours students, but may also be true of students enrolled in higher-level degrees, or even supervisors. It may not be till several years after graduation that most of us integrate all the various topics we studied as undergraduates into a coherent model of computing-as-a-whole, which supports and helps us direct our research endeavours.

Too often, research students also have no idea of the real reasons behind the development of their particular systems. You aren't just slaves implementing something your supervisor has dreamt up - no matter how much it might seem that way sometimes. Underlying your work is a *thesis* - a belief, if you like, though that implies a rather less reasoned commitment to an idea than is appropriate for an academic researcher - that something is true. The something may only be "it's worth developing a system like this" or "there are interesting trends to be discovered by summarising the research of this group of people", or it may be more fundamental: "it should be possible to solve this problem by an elegant new algorithm based on ...". In any case, you need to make yourself aware of the thesis, and develop the ability to express it, and to relate your research decisions back to it.

---

* Gestalt: Perceived organised whole that is more than the sum of its parts, e.g. a melody as distinct from the separate notes within it. *Gestalt* psychologists hold that perceptions, reactions, etc., are *Gestalts*. From the German for form or shape.

Further, there will almost certainly be related, lower-level, theses which control relatively minor decisions you have made during the research, or which are supported by the intermediate discoveries you've made during the research. These are the most notable omissions from students' reports on their research. If you don't justify intermediate-level decisions you've made, or you don't draw the intermediate-level conclusions resulting from their discoveries, then your report will lack coherence, and may project an image of doubt ("I'm not sure why I did this") or arrogance ("This is too obvious to need explaining"), or even, sometimes, both. If you do provide justification for your approach, your reader will feel that you have a good grasp of the subject, and will feel more inclined to believe you if you slip up and make an occasional bald statement where there should be an argument.

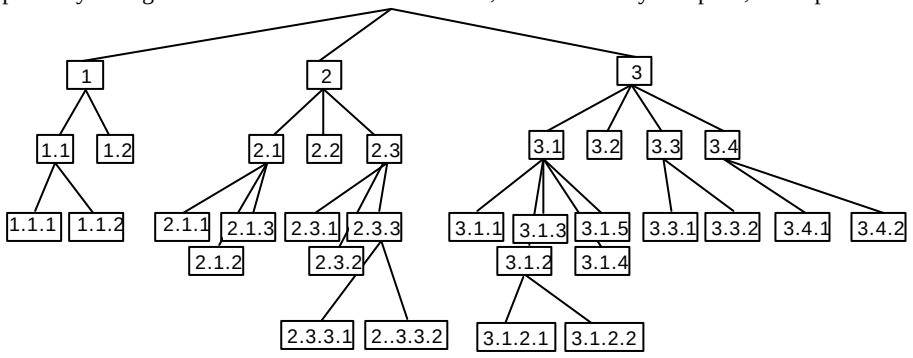
Within the overall scheme of the report, students most often omit this contextual discussion from the *problem specification*, *discussion of other approaches* and *conclusions* sections. There are two approaches to circumventing this tendency. The first is to improve your gestalt throughout the project, so that it is as nearly complete as possible when you come to explaining yourself. You can do so in many ways, such as reading lots of relevant literature*, running seminars with question time, and conducting regular discussions with supervisors and other students. The second approach is to look over your documentation critically, preferably with the help of a proof-reader, ensuring that you have justified all its assertions, that all the justifications actually support some assertion, and that you're neither teaching grandmothers to suck eggs nor requiring your readers to have completed a Ph.D. in the area.

Navigation within the report is your responsibility

Let us return to the general problem of mapping a *gestalt* onto linear text. As I have intimated, this is largely a problem of making real links apparent and of making links seem to exist where there are none. Linking the document at the highest level is fairly straightforward; it is achieved by deciding upon the reasonably standardised chapter sequence within the report.

It's within chapters, however, that problems arise. Computing students in particular - aided by the eagerness of implementers of word processors to support easy-to-implement features - structure their document as a multi-level tree, and expect the reader to use multi-level decimal indexing for navigation This can certainly help the *author* to be sure that everything has been included, and that the presentation has a logical sequence. However, it doesn't give nearly the same quality of information to the reader, so *it is the wrong approach.*

Try to think of it from the reader's point of view. She or he doesn't have your *gestalt* appreciation of the ocean of material, and is relying on you to be the pilot. Imagine then if the primary navigation tool is a structure like this, not unusually complex, example:



* particularly recent research reported in current journals, in preference to the "old news" from any journals which your supervisors may have read in preparing the project, and summaries of even older news contained in textbooks.

Of course, unless the sections are very small - which brings their necessity into question - the reader can only see one or two on a page at a time. How, then, can this complex numeric information possibly elucidate the *context* of the *current point* in your argument?

Remembering that the navigation information should support the argument, not *vice versa*, try to imagine the reader starting to read Section 3.4. If the author has structured the document purely by top-down analysis, the context of subsection 3.4 will be in subsections 3.1, 3.2, and 3.3; the subsections *at the same level*. The reader will have to reconstruct the sequence involved in those subsections while suppressing all the lower-level discussion which she or he has just read. Without extra navigational information supplied by the author, who knows what relates to what, this is not an easy task. It is *your* responsibility as author to perform it, not the reader's.

Decimalised numbering systems are intrusive, and say nothing about the material under discussion. However, they allow you to pretend that navigation is the reader's responsibility, when it is your responsibility to perform it as subtly and invisibly as possible. You must, by devising seamless connections between the sections of your argument, perpetrate the falsehood that the argument is a sequence of related points, each placed in a position inevitably following the previous one. To achieve this neat interrelationship totally, without using heavy-handed lists of points to be covered, and recapitulations of earlier points, is a first step in achieving a consummate mastery of *technical* writing.

Here's an example of navigational writing in which the author unobtrusively linked a pair of points with little in common:

> Thus a major problem in implementing the translator had been solved. There remains one other problem to be tackled, however. It is ...

We can see that one topic has been completed, that another is about to be started, and the minimal extent of the relationship between them. Thus the linking words both emphasise the problems' common property - the fact that both are problems - and alert the reader *not* to search for deeper links between them. Further, the reader assimilates this information easily without distraction from the important part of the discussion.

As I mentioned earlier, word processors' navigational provisions can be more harmful than helpful, particularly the decimalised numbering systems. However, intelligently used, outlining can be a useful aid in organising your thoughts.

Probably the most difficult part of the mapping between the *gestalt* and the text occurs after you have made the major "strategic" decisions about the contents of chapters and their major sections. You must then compose and sequence individual paragraphs. The advice which teachers traditionally give to essay-writers at school may help: allocate one paragraph to each thought. Start with a sentence which expresses that thought, or leads naturally to it. The rest of the paragraph examines, support, contradicts or extends that thought.

Outliners support this structure admirably. They allow you to take a first stab at the sequence of paragraphs within a section (let us say) by writing that section as though *it* were a single paragraph, comprising sentences each of which you will later expand to a paragraph. Thus you can summarise your argument in a series of terse and unsupported statements, each destined ultimately to become a subsection heading, or the first sentence in a paragraph. Subsequently, you can flesh out this skeleton. In doing so, you will not merely expand on the material in the introductory statements, but also use it to help you find the correct place in the overall argument for the random (but relevant) thoughts that occur as you write. That is, if an idea occurs to you as you write, don't automatically include it in the current location; look for the pre-written heading under which it best fits, and put it there. Then you won't forget it, but it will be in context.

How do I avoid common writing errors?

Enrolment in a technically oriented degree does not involve much assessment of your ability to write English-that-matters-as-English. Indeed, much of your previous course assessment has required no real writing at all, being largely concerned with formal languages such as programming languages, definition-by-diagram or various sorts of mathematical formalisms. Consequently, you are far less practised at writing than your colleagues enrolled in, say, History or Philosophy. Now you face the challenge of writing a document in which you must augment your undoubted clarity of *thought* - after all, you have reached a reasonably elevated level in your subject area - with clarity of *writing*.

Your work is not going to impress markers and future researchers if the language used to describe it is full of minor errors. You may feel the errors described below are *too* minor to interrupt a reader's flow of understanding and divert her or his attention away from the effort to reconstruct your mental model. However, many readers find such errors irritating in the extreme, and your writing is for the readers' benefit, not yours. Forcing them to quarry nuggets of meaning out of a lode of rotten English would be arrogant, if deliberate, and doltish otherwise.

So let's look at a list of misunderstandings which have caused many of the errors which I particularly remember finding in research reports. I emphasise that they are my own, personal, antipathies. Some readers will have different ones; other readers may consider these to be excessively pedantic, or may not even care. I include them mainly to increase your awareness of the English syntax and style, and of the desirability of paying some attention to them.

The items occur in no particular order. Like Topsy, the list "just growed". In a couple of cases I've presented some relevant advice, but not an actual error, and some points of more general advice have crept in as well.

Be consistent in your use of tenses

If you're describing what *happened* (like design decisions), use the past tense. If you're describing something that *is* (like the capabilities of your system), use the present tense. If you're defining something, use the present tense; a definition belongs to no particular time. Tense usage should be consistent throughout the document.

Maintain syntactic parallelism between repeated parts of a sentence

Consider the following sentence:

> This operation is either called when a click event occurs over a primary object or as part of another operation.

The sentence has two subordinate sections, under the control, as it were, of the either-or construction. It is essentially a contraction of two sentences outlining a pair of alternatives, something like:

> This operation may be called when a click event occurs over a primary object. Alternatively*, this operation may be called as part of another operation.

The original sentence has a beginning and two alternative endings. We should be able to place either of the endings after the beginning, to make a syntactically valid sentence, but look at what happens if we try this:

> This operation is called when a click event occurs over a primary object (OK)
>
> This operation is as part of another operation. (not OK)

The items following the "either" and the "or" respectively are different syntactically, and make nonsense when extracted and treated separately. The reason becomes clearer if we parenthesise the relevant sections of the sentence:

> This operation is **either** (called when a click event occurs over a primary object) **or** (as part of another operation).

Their different syntactic natures are now immediately apparent, but how do we solve the problem? We shift the words around, to achieve *syntactic parallelism*:

> This operation is called **either** (when a click event occurs over a primary object) **or** (as part of another operation)

which is correct, or:

> This operation is **either** (called when a click event occurs over a primary object) **or** (called as part of another operation)

which is better. (Why?)

Maintain syntactic parallelism between parts of a list

When you find a whole group of similar things to say, you may find that a list is the appropriate way to encapsulate them without needless repetition. If the reader is to assimilate the list easily, its structure must not interrupt the flow of the writing. Although it may appear to be a separate entity, like an illustration, you can, and should, integrate it more closely into the structure of your narrative. In particular, all the components of the list should be the same part of speech, and should each be the continuation of the preceding phrase. Consider this example:

> In dealing with this system, we have found it useful to categorise the user-community into beginners, practitioners, and masters on the basis of their level of competence. They have the following characteristics:
>
> > The absolute beginner
> >
> > The ability to accomplish simple tasks without recourse to reference manuals
> >
> > Total familiarity with all aspects of the system

* NOT "alternately". Look up "alternately" and "alternatively" in a dictionary if you're not sure of the distinction. Look up lots of things in a dictionary, as a matter of principle.

It's not too hard to see what the writer is saying, but the use of a syntactically different construct for each of the three categories causes a moment of syntactic confusion, and hinders the integration of the information into the reader's mental model. Further, each of the categories is the continuation of a sentence, but a completed sentence precedes the list. I've rewritten the example, integrating the introduction and the list, and making the elements syntactically identical:

> In dealing with this system, we have found it useful to divide the user-community into three categories based on levels of competence. The beginner's, practitioner's, and master's levels are occupied respectively by users who can:
>
> > perform simple tasks with outside tuition;
> >
> > perform tasks of medium complexity untutored, but with recourse to reference manuals;
> >
> > perform any task which the system can handle.

You will have noticed that I've also separated the list elements with semicolons, and omitted the introductory capital letters. If they included complete sentences, I'd have avoided making the elements into continuations of an introductory sentence, and would have ended them with full stops.

Choose your voice with care

You can write using the passive voice ("the program *was written* in two weeks") or the active voice ("*I wrote* the program in two weeks").

It is hard to decide whether you should write in the passive voice or the active voice. On one hand, there is a large body of entrenched opinion - derived originally, I believe, from Royal Society pronouncements in the 18th Century on the correct way to write - which holds that the passive voice is the correct way to depersonalise scientific statements, and reflects their universal validity.

This was an appropriate direction to give writers of the time, who had no prior examples of good scientific writing to use as models. Further, at the time, the Royal Society thought that scientists really could make impersonal, universally true statements about their subject matter, the nature of the universe. We more technically oriented writers inherited this scientific style without giving much thought to it. The resulting depersonalised writing has a spurious ring of authority, but saves the reader from having to wade through a tedious succession of statements with an "I did this", "I did that", "I think that...." construction.

On the other hand, modern developments in science (uncertainty theory, and all its consequences) have led scientists to believe that the nature of the universe is not as fixed and "universal" as had previously been assumed. There is a growing recognition of that most uncomfortable of notions, that the way we *look* at things influences the way things *are*. Consequently, there has been a trend to more personally oriented writing, even in mainline scientific disciplines like physics. Similarly, the spurious assumption of authority that comes from passive-voice writing (the "it was shown that" and "the system has been designed to achieve the following benefits..." sorts of statement) is coming into question. If *you* have shown something or designed an approach, why aggrandise it with passive construction? Don't you personally stand behind your demonstration or design?

I seem to be tending towards a suggestion that you pepper your report with first-person singular pronouns ("I"s and "me"s). However, as I pointed out above, this can make for a style

which is just as tedious as the most circumlocutory passive-voice writing ("the opinion has been expressed elsewhere by the present author that it is possibleto..."). Unfortunately, whereas a relentlessly passive document hides its writer's identity behind a screen of authority, a relentlessly first-person singular document forces its writer's identity into the reader's consciousness.

There doesn't really seem to be a totally acceptable middle ground. The best suggestion I have heard is to write in the first person plural, ("we designed the system to achieve the following benefits...") even where you are describing something you developed entirely on your own. This works fine, except when it really matters that *you* did the work - when your document presents your own, personal, research for marking.

So, after all that discussion of the *pro*s and *con*s of passive-voice writing, I don't offer any absolute advice. Be aware that passive-voice writing can lead to circumlocution and an appearance of false authority, that first-person singular active-voice writing can be tedious to read, and that first-person plural active-voice writing can reduce your apparent personal contribution to the development.

Maybe the best advice is pragmatic: many academics have been trained in the use of the passive voice in their writing, and they are firmly convinced that it *must* be used, so it's probably unwise to choose another approach without clearing it with your supervisor first.

Finally, you'll probably have noticed that I use the passive a lot. I try not to, but it was beaten into me when I was trained as a scientist, and I just can't help myself.

Commas matter

Generally, when you interrupt the flow of a sentence, you should insert a comma. Many inexperienced writers omit them, but they are often essential to the sense of the writing.

> When a procedure or function is called the actual parameters must match the formal parameters.

When you parse the sentence above- and you automatically parse everything you read- you have to contend with the fact that the object of the verb "called" could be the phrase "the actual parameters". It's not till you reach the word "must" and are forced to backtrack that you discover that "called" is being used intransitively (without an object), and means "invoked", and not "named". A comma after "called" would prevent the confusion.

Consider the following, more complex, example

> The resulting variable, **pixel**, is 24 bits wide where the red component is contained in the first eight bits.

Here, the writer tells us two things: first, that there is a variable called **pixel**; second, that in a particular circumstance, "where the red component is contained in the first 8 bits", it is 24 bits wide. The reader would be justified in inferring that, in circumstances where the red component is *not* contained in the first eight bits, **pixel** is not 24 bits wide.

Alternatively (and because we happen to know that colours are often stored as a 24-bit value with eight bits allocated to each of the red, green and blue components of the colour, we would probably infer that this is what the student meant, in spite of the misleading punctuation):

> The resulting variable, **pixel**, is 24 bits wide, and the red component is contained in the first eight bits.

Now the information conveyed to us is: first, that a variable exists and is called **pixel**; second, *that it is* 24 bits wide; third, that of the 24 bits, the first 8 contain the red component.

Finally (and this is what the student actually wrote):

> The resulting variable **pixel** is 24 bits wide where the red component is contained in the first eight bits.

In this example, because there is no comma between "variable" and "**pixel**", "variable" is just an adjective qualifying "**pixel**". To put it another way, it looks as though the student is drawing a distinction here between a *variable **pixel*** and a *constant **pixel***, which is presumably referred to elsewhere in the document. In fact, the word **pixel** is really meant to be just a parenthetical insertion in the sentence. The *syntax* of the sentence does not require it at all. English allows a variety of ways of inserting parenthetical words, phrases, and clauses. They can be inserted between commas, as this one is, between dashes - as this one is - or between brackets (as this one is). Commas are the most frequently used of the three alternatives.

Now, when proofreading the thesis that contained the example above, I was able to determine from context that the second of the three interpretations is the one that the student intended. So why worry about putting in commas at all? Because I had to divert my attention from the task at hand - building up a mental model of a data structure for representing pictures - to line up the possible interpretations and reject those that were probably incorrect. If your reader has to do this too often, she or he will conclude that your writing portrays the subject poorly, and, therefore, that you don't really understand it very well. The first part of the conclusion is valid, the second part possibly not. In any case, it's your responsibility to make your text as limpid as possible.

Articles and restrictors

The words *the* and *a* and *an* are called articles in grammar. They can be confusing, especially if your first language isn't English. When I tried to write a description of how they fit into English, I discovered that they're hard to categorise if you treat them on their own. However, if you recognise that they're part of a more general class of restrictors, it's possible to develop a more general classification. That's what I've done here. I'm grateful to Arianna Berardi-Wiltshire (a Massey University linguistics specialist) for convincing me to be a little more systematic about this classification than I was at first. Her advice has strengthened the description considerably; however, any faults that remain are entirely my own responsibility!

Words that restrict nouns: *A* car, *the* car, *some* car, or (just) car?

When we write down a noun[1] on its own, without any preceding restrictive word, it functions primarily as the name for a whole class of things, rather than an individual item. *Cow* doesn't mean *Daisy, the cow with the crumpled horn*, but the class *cattle beast, horned, female, adult*. So if we want to talk about a property of cows in general, we can write

[1] Strictly speaking, I ought to be talking about noun phrases. In the sentence *The car with a red roof drove past*, the subject of the sentence (i.e., the thing that drove past) is the noun phrase *the car with a red roof*, not the simple noun *the car*. That's a noun phrase, but it's easier to talk about simple nouns, so I will.

Cows are second-rate tunnellers.

Of course, we don't always want to make generalisations about a complete class of items; sometimes we want to restrict our discussion to a subset of the class - either a subset that we haven't talked about previously or a subset that we have talked about before. English uses a set of modifier words to restrict nouns so that they refer to various subsets of the class, rather than the class as a whole.

So whenever you write a noun, ask yourself the following questions:

Do I want to write about the properties or behaviour of a whole class?
If not, what prefix word will restrict it to the correct subset of the class?

We'll start with nouns used in their most general sense[2], that is, as the names for complete classes. Then we'll see how different restrictive modifiers can turn the class name into a name for successively smaller and more closely defined subsets of the class, until it refers only to a single, previously identified, member of a whole class.

We've already seen an example of a noun being used to refer to the behaviour or properties of a whole a class of things. In that situation we use a noun with a plural ending and we don't put any restrictive modifier in front of it. For example, if we want to discuss the behaviour of cars in general, we'd write

Cars should stop at traffic lights when the lights are red.

The word *cars* has a plural ending and it has no restrictive prefix, so we know that the statement applies to all cars. Conversely, although the word *cars* looks like a plural, the lack of a restrictive tells us that it's really a classname, so the statement also applies to all members of the class, including single cars. That is, even a single car must stop when the light is red.

Unidentified sets of class members

The first type of restriction converts a noun from a classname to a name for a set of class members that has not previously been discussed. Some restrictive modifiers restrict a class lightly or not at all

All cars, most cars, many cars, lots of cars…

some restrict a class to a vague extent

Some cars, a number of cars…

some restrict a class a lot

A few cars, few cars, hardly any cars, no cars…

and some restrict a class to a subset of an exact size

One car, three cars, a car…

[2] Well, not quite. In this context, there are two types of noun…
Countable nouns are used as names for things that are measured by number; that is, discrete individual items like cars, commas, and cows. *Non-countable* nouns are used as names for "stuff" that's measured by quantity; that is, continuous stuff that has no obviousboundaries, like milk and putty. We'll talk about the rules for countable nouns first. Then we'll deal with non-countable nouns. However, you'll be pleased to hear that non-countable nouns behave very similarly to countable nouns.

We use these restrictive modifiers when we're writing about a subset that has not previously been discussed.

The noun we're restricting has a plural or singular ending according to whether we're writing about a subset that contains multiple or single items.

A cow dug a tunnel under the fence singular form

Some cows escaped through the tunnel plural form

Identified sets of class members

The first time a reader reads about an unidentified subset of a class, she or he creates a mental model of that subset. Once the model has been created, the subset is said to have been identified, and when you want to refer to it again, you need to tell the reader that you're referring to the previously identified subset of the class. You do that by prefixing the classname with the definite article (the word *the*). *The* is both singular and plural; we refer to *the cow* and *the cows*.

Classes with only one member

A class that has only one member is an interesting special case. When everyone knows that it can contain different members - but only one a time - we restrict the class name to its single member by automatically giving it a definite article (the). For example, we can write about *the price of milk, the German Prime Minister, the average height of buildings in France, the type of footwear used by cowboys* without ever having mentioned milk, Prime Ministers, buildings, or cowboys, anywhere else in our document. And if our family owns a cow, we can even write *the damn cow's escaped again*, because we know exactly which cow is meant without having to identify it explicitly (we should never have bought a cow with a crumpled horn; they're always unnaturally talented tunnellers).

On the other hand, if the class and its only member are indistinguishable, we don't need to perform any restriction, so we don't bother with a definite article. Proper names fall into this category. We say *Helmut Kohl was in Berlin on Monday*, but *The German Prime Minister was in Berlin on Monday*. This can give rise to interesting situations:

"Are you Charles Dickens?"

"Yes."

"Can you sign this copy of your book?"

"I didn't write that book."

"But it's by Charles Dickens."

"Not by me."

*"Oh, so you're not **the** Charles Dickens."*

"Yes, yes I am." For this speaker, his existence and his ego prevent the class of Charles Dickenses from having any other member worth speaking about.

The following sentences should illustrate the various types of restriction that can be placed on a noun:

If we want to write about…	***then we might say…***
the class in general, with no restrictions	***Cars*** *must stop at the traffic lights when the lights are red*
a subset of the whole class, not identifiable from previous discussion	***Most cars*** *stop at red lights, but* ***some cars*** *just drive straight through. I saw* ***a car*** *crash a red light yesterday.*
a subset of the whole class (one item or several items) that is identifiable from previous discussion	***The car*** *that hit me had a green roof.* **or…** ***The cars*** *all had green roofs*

Restricting non-countable stuff

The discussion so far has dealt with *things*; things that can be counted. Now we're going to move on to *stuff*; stuff that doesn't have well-defined boundaries, like milk and water. Formally, the names of things are *countable nouns*, and the names of stuff are *non-countable nouns*. I prefer things and stuff. We'll find that the rules for stuff are similar to the rules for things, except that you don't have to worry about singulars and plurals, because stuff doesn't have plurals…

When we refer to the behaviour and properties of the stuff as a class, we don't use a restrictive (just like things). Here's an example that illustrates the behaviour of both types of noun

Milk is produced by cows.

As the names of general classes, neither *milk* nor *cow* needs a restrictor. *Milk* is the name of a non-countable stuff, so it doesn't have a plural, but *cow* is the name of a class of countable things, so it does have does have a plural.

When we refer to a previously unidentified amount of stuff, we use a restrictive that doesn't imply counting, (*some milk*, or *a little milk, a cupful of milk, a gallon of milk*), but we may not use a restrictive that implies counting (*three milks*, *many milks*, and never *a milk*[3]).

When we refer to some stuff that's been referred to previously, we restrict it by using the definite article (*the milk had soured).*

[3] Well, hardly ever. See the next section…

We can add some more sentences to the table we drew up before:

If we want to talk about…	***then if it's a class of things, we might say…***	***and if it's a class of stuff, we might say…***
the class in general, with no restrictions	***Cars*** *must stop at the traffic lights when the lights are red*	***Milk*** *comes in udders, bottles, and waxed cardboard boxes*
a subset of the whole class, not identifiable from previous discussion	***Most cars*** *stop at red lights, but* ***some cars*** *just drive straight through. I saw* ***a car*** *crash a red light yesterday.*	***Some milk*** *has been spilt on the carpet* **or…** *I bought* ***a gallon of milk***
a subset of the whole class, identifiable from previous discussion	***The car*** *that hit me had a green roof.* or… ***The cars*** *all had green roofs*	***The milk*** *had turned green*

A horrible exception

Much academic progress relies upon detecting fine distinctions between things that people have never previously bothered to distinguish. Sometimes those things are classes of non-countable things. For example, we might have to write a survey of different types of milk (milk from grain-fed cows, milk from potato-fed cows, milk from free-range cows, milk from cows that smoke dope, etc.) In those circumstances, the non-countable material becomes a countable item and we could justifiably write about *a milk*.

> *It has been shown that 58% of respondents prefer a milk with a calcium content exceeding x%.*

Or even

> *In a recent survey, European subjects preferred the milks from grain-fed cows and potato-fed cows by a margin of 7%, with a standard deviation of 2.3%.*[4]

I include this appalling locution because, as an academic writer, you are more likely than most other writers to have to use it.

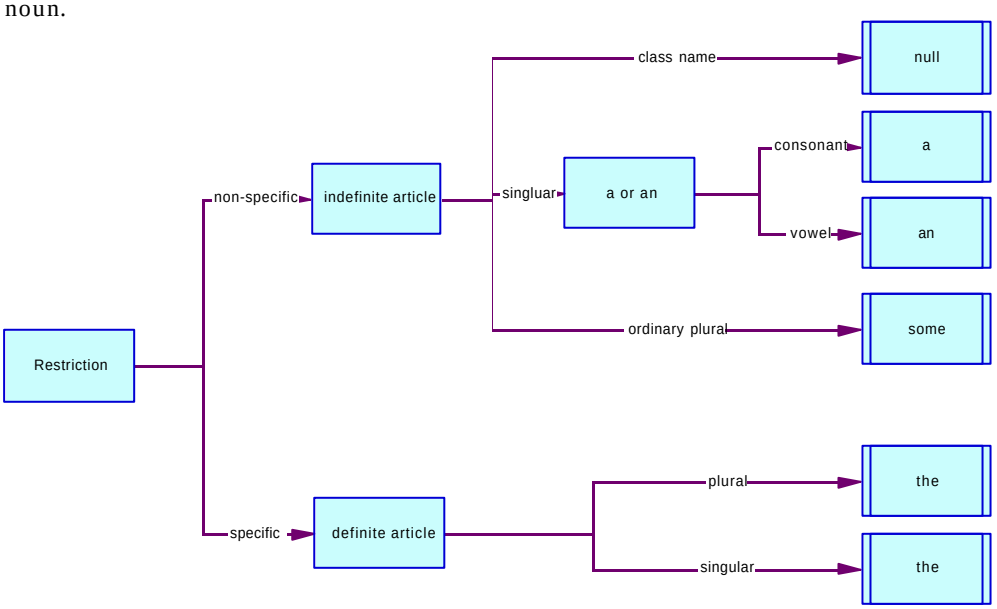
Adjectives

Where do adjectives fit into this scheme? They don't.

Adjectives may seem to define a subset of a class, like restrictives, but they're in a different (you guessed it) class. We seem to treat *yellow cows* as a new class (and possibly as a biological aberration), rather than as a restriction on the class of cows. It seems a subtle

[4] All right, you've caught me out. I'm making these statistics up. I neither know nor care whether Europeans prefer milk from grain-fed cows. And I think the standard deviation might be wrong.

distinction, but it explains why we can write *some yellow cows* or *the yellow cows*, but we could never write *yellow some cows* or *yellow the cows*. All clear now?
Here's a diagram that might help you to choose which, if any, restrictors to put before a noun.



Although the diagram looks daunting, it's not really very complex. You'll probably find that you understand at least three quarters of it without difficulty. But if English is not your first language, you may need to spend a little time revising the distinction between conventional plurals (where we're talking about a group of specific items) and class names (where we're talking about the properties and behaviour possessed by a whole class of items). These are the top and bottom items at the right hand edge of the diagram.

Exclamation marks

Restrict yourself to one exclamation mark. Yes, that's right; one in the whole report.

"E.g."s, "i.e."s and apostrophes

Many students don't seem to be aware of the distinction between "e.g." and "i.e." They are not interchangeable. "E.g." is an abbreviation (hence the full stops) of the Latin phrase *exempli gratia*, which means (though this isn't its literal translation) "for example". You should use it when the item(s) following the "e.g." *illustrate*, but do not recapitulate the point made before the "e.g.".

"I.e." is an abbreviation of the Latin phrase *id est*, which means "that is". You should use it when you are restating a point in different words, i.e., when your first expression of something needs reiteration. E.g., if you wish to use a technical term which you think the reader may not understand, you could follow the term with "i.e." and then a definition.

If you use "e.g." or "i.e." excessively, (as I have in the previous paragraph, for illustrative purposes) you'll distract the reader from the point you're trying to make. It's then probably better to write the phrases out in full (in English).

Abbreviations, except some of the most common,are followed by a full stop. When you pluralise them, you add the letter *s*. In almost every case, there's no need for an apostrophe. There is a growing tendency, in informal writing such as shop signs[5], for words with "funny" endings, like words ending in a vowel, such as *piano*, to be pluralised with an apostrophe before the final *s*. I.e., the plural of *piano* would be written as *piano's*. This is incorrect. Don't do it.

Apostrophes are often misused. They may be used to indicate possessives, and to indicate that letters have been omitted from the middle of a word (including where the word is a contraction of two words). So correct usage gives us:

The computer's down	(contraction of "computer is")
The computer's disk is down	(possessive, meaning "the diskof the computer")

In generating possessives, there is a rule which is, with two exceptions, invariably successful: put the apostrophe after the thing that does the owning. So:

The user's input	(only one user)
The users' terminal room	(lots of users)
The plus' visual representation	(the visual representation of a plus)

What are the exceptions?

Words ending in *s* can *look* wrong even though correctly apostrophised. Sometimes, therefore, we add an extras. Thus *the plus' visual representation*, above, might be written *the plus's visual representation*. In either case, the plural would be *the pluses' visual representations*.

It's means *it is*, or *it has*. It should *never* be used to indicate possession of something by an inanimate object. Thus:

The computer is down. *It's power supply exploded*.	is incorrect, whereas
The computer is down. *It's had power supply problems*	is correct. And true, too, too often.

Similarly, don't ever use *her's*, *their's*, *our's*, or *your's*. They don't exist in English.

Americanised spellings of English

Although New Zealand spelling follows British, rather than American, models, there are some technical terms in which - barring a few intransigents - the whole world has agreed to follow American usage. Principal amongst these are: "Program" (when a computer program, and not, say, a conference programme, is being discussed) and "disk", when a computer's secondary storage device, rather than, say, an old-fashioned LP, (a disc) is being discussed.

[5] This occurs so often that this incorrect use of the apostrophe is sometime referred to as “the grocer’s apostrophe.” See, for example, the hilarious diatribe about grammar called “Eats, Shoots and Leaves.”

Use gender-inclusive language

It is now widely accepted that the words *he*, *him*, and *his* are subconsciously interpreted by most readers as references to males only. In the 1996 BBC Reith Lectures, unattributed University research into this matter was described. It was found that on hearing a sentence such as: *If you see a botanist in a field, you can assume that he is working* a large majority of people would assume that the biologist was male, whereas less than 50% assumed that the biologist was male if the sentence were rephrased as *A biologist in a field is probably working*.

Don't use sexist language in your report. If you are referring to a particular person, or to something associated with females or males only, then use gender-specific pronouns. Otherwise use "she or he", "her or him", or "hers or his". The constructions you end up with can sometimes seem clumsy, but they are preferable to being sexist.

Colloquialisms are out of place in technical writing

Avoid colloquialisms and attempts at humour (unless you're *very* sure of the maturity of your wit) in your text. Like so many of the other examples of bad writing practice, they are not, in themselves, or in the right context, a bad thing, but when you are trying to convey information in the clearest possible way, you want to keep the full attention of your reader (whose first language need not be English, and whom your subtle humour may confuse) all the time.

Don't claim to have implemented what you have only planned

Make it clear what you have actually implemented (and tested, and found to work) and what you have only specified. It's very easy to get carried away in the description of a system's specification, and make statements like *The system generates a complete set of...*, or *Error-recovery procedures act thus...*, when in fact these are only part of your (well organised) planning, and the system doesn't really do them at all. Yet.

Make it *quite* clear what exists, and what doesn't. An assessor who notices inconsistencies in this area can easily mistake them for deliberate attempts to mislead, and withhold marks as a result.

Hardly anyone reads section headings

Well, some people do, but they are in a minority. By all means put useful information - whole statements for preference - into section headings, but if you want to be sure it's read, repeat it in the text.

Justifications need conclusions

Make the point of the paragraph clearly - don't just state the supporting arguments and leave it up to the reader to guess what they support. Sounds silly? Consider the following two examples from an Honours report:

EXAMPLE 1

> When there is more than one protocol currently in use, a need to compare the contemporary ones arises even if the superiority of one protocol can be firmly established by a cursory

> overview. This is because a comparison can help the creation of a document outlining the specifications for the constructions (sic) of gateways between them. Even in communities that intend to make a smooth transition to the superior protocol, these gateways or mail relays can ensure a smooth transition service.

In Example 1, the point of the paragraph is stated in the first sentence. Supporting reasons then follow, and are identified, so that the whole is well integrated. The right approach, in fact.

EXAMPLE 2

> As can be seen, the protocol information is buried deeply within the DECnet chatter and is very hard to filter out. Besides, the unreliability of network monitoring software further complicates the issue by not guaranteeing to trap every packet that goes by it. In particular, it can be noticed that the confirmation for the Addressee list that should be sent by the slave (0218) to the Master (0718) is missing.

The first sentence of EXAMPLE 2 *begins* with a supporting reason ("the protocol information is buried..."), then continues with the real conclusion ("is very hard to filter out"), in exactly the same syntactic environment *as if it were just another reason*, and then reinforces this invalid presumption by starting the next sentence with "Besides", which would normally connect two sentences of justification, not a conclusion and a justification.

But maybe Example 2 doesn't really seem too bad to you. After all, it's quite possible to sort out the real point of the paragraph. However, if you have to read things twice to sort out the deep meaning, your smooth flow of understanding is disrupted. Having to do this two or three times in succession can quite prevent you from deriving any useful, connected model of the material in that part of the report.

The example should seem better when reasons and conclusions are properly separated:

> The information is hard to extract for two reasons. First, it has to be filtered out from the DECnet chatter. Second, the network monitoring software is unreliable and cannot be guaranteed to trap every passing packet. For example, confirmation for the Addressee list that should be sent by the slave (0218) to the Master (0718) is missing.

Notice that the "further complicates the issue" clause has been done away with. In the original, it looked like meaningful information, but seems superfluous in the rewritten version. Watch out for floss that can be eliminated when you improve a paragraph's structure.

Report length and condensation

An initial project specification which is vaguely defined, with many subsequent ramifications, is likely to result in a larger report than a tight specification followed by a project which runs smoothly to a successful conclusion, so it is impossible to prescribe absolute limits for report sizes. And reports are not assessed on weight, contrary to some theories.

Nevertheless, one might find a marker wondering about the sufficiency of research which results in an Honours report of fewer than, say, 20 to 30 pages (approximately 10000 to 15000 words)[6]. Someone who has completed a substantial piece of research will rarely be able to

[6] Master's thesis, 100 pages (55000 words); Ph.D. thesis, 150 pages (80000 words)

summarise its background, content and conclusions in less space than that (due allowance being made for the level of the project).

Similarly, Honours reports with more than 100 pages (55000 words[7]) *may* be reporting a wide-ranging project with awesome conclusions, but are more often full of redundant or irrelevant information, or just plain bad writing. Many markers are just as prepared to mark students down for not knowing what to exclude as for having nothing to say.

A document such as this can't do much about an insufficiency of basic research which leads to too short a report. That is a problem for the student and the supervisor to deal with. However, reasonable amounts of research often get written up in too many words, particularly when the writer is trying to develop an impressive "academic" style. Consider these two examples:

> *all* its aspects are at the forefront of your mind, and the interrelationships between the important points become very clear to you
>
> *all* its aspects are at the forefront of your mind, and their interrelationships are clear

The second says the same as the first, rather better, and with about 30% fewer words. That's a reduction of a 150-page thesis to a 100-page thesis. Most first-draft writing can be condensed in this way by a similar amount. Notice that some irrelevant things have been omitted from the second version. The aspects (or points) being interrelated are no longer just the *important* ones. The interrelationships are no longer *becoming very clear to you*; they are just *clear*. In addition, by omitting *points* (a synonym for *aspects)*, we no longer confuse the reader by forcing him or her to decide whether *aspects* and *points* are really identical, and the writer was just trying to avoid repeating a word within a sentence, or whether they are subtly but significantly different.

None of these changes is individually very significant[8]. However, by treating an entire document thus, you can trim it to the bone; if you have also included all relevant navigation information (see above), and made sure that all your arguments lead to justified conclusions, then the true beauty of your research will shine through, and you can be sure that the mark will reflect that.

Proofreading, and dealing with proofreaders' comments

First; should you have your report proofread? Yes. A spelling check isn't nearly good enough as a check on your writing. It won't even detect all the spelling errors. There are many words which are commonly misspelt to form other valid words: there/their, for/four/fore, you're/your, and so (sew?) on. For example, Word's grammar checker (which incorporates a spelling checker) accepts the following "statements" without complaint:

> **Wen wee plaid their, wee one bye too gaols two won. Hour captains' suns putt hour to inn (won purr sun).**

Good proofreaders are rarer than hen's teeth. You want someone who can pick out and note the trivial errors (spelling, punctuation, grammar, unnecessary colloquialisms and so on) and still have brainpower left to criticise the overall flow of the writing.

---

[7] Masters' theses, 200 pages (110000 words); Ph.D. theses, 250 pages (130000 words)

[8] I first wrote "is a very significant one". "One" in that context can generally be dispensed with, to make the text leaner.

Treasure such a person when you find one. She or he can help you enormously, but you have to be prepared to handle criticism. Most of our (Computer Scientists') writing is poor. Most of it could be improved enormously. So most of it deserves trenchant criticism, and that's going to hurt. We put a lot of effort into our writing, even bad writing, and harsh words about it are personally upsetting.

So, don't ever take a proof reader's criticisms to heart. Adopt the attitude that you have the task of building a delicate structure by remote control inside the reader's head. The text you write is, if you like, a program for controlling this process. Any difficulties the reader has in interpreting it will lead to errors in the resulting model. This applies to any parts that the reader has to re-read because they're unnecessarily complex, or poorly ordered, or inadequately related to the overall context of the discussion. It also applies to parts that cause the reader's attention to be diverted away from model-building to deal with bad grammar, incorrect terminology, or flippant remarks (particularly comments suggesting the writer's lack of competence - a curiously common occurrence in students' work).

The purpose of employing a proof reader is to subject your writing to close scrutiny by someone who is sensitive to all possible distractions from the message. She or he needs to point out the stylistic, or grammatical, or typographical causes of these distractions. Then you can remove them so that they won't trouble future readers. This is the key to your relationship with the proof-reader; if she or he is distracted from the *point* of the writing by the *nature* of the writing (no matter how attached you may feel to the particular method of expressing the point), then later readers will probably be distracted too. Think long and hard before rejecting the proof-reader's suggested improvement. By doing so, you risk misleading your target readers. Provided that the proof-reader's suggested improvement actually means what you intended, restrain your pride and accept the suggestion.

Diagrams and program listings

Diagrams are a great way of grabbing a reader's attention. A reader, even if only mildly interested, will be intrigued by a clear, visually well organised diagram, will read the legend, and if that is interesting too, will go on to investigate the textual description.

However, you are more concerned with dedicated readers, for whom the text is the most important "way in" to your document. Do include diagrams and program listings to illustrate your points, particularly if they are complex, but don't forget to use the text to direct the reader's attention to the important features of the inclusion. This doesn't just mean including a reference "(see Fig. 3.7)" at the end of a section to which the diagram pertains; if you don't make *specific comments* relating to *particular components* of a complex figure, your reader is quite likely to ignore it completely. Then, instead of acting as a thousand-word addition to your argument, it acts as a thousand-word omission.

Bons mots

Most of us are not literary figures of any great stature, and our prose is rough. In spite of that, many of us like to include catchy phrases in our reports. Now, these may be apposite and genuinely illuminating, but even when they are, they often serve to disrupt the flow of the text. And the more highly polished they are, the more they expose the roughness of the remainder.

How do I refer to The Literature?

Information which is more detailed than could be obtained from a standard text on the subject, and which is derived from others' work should always be referenced. The format for this is quite rigid, to avoid confusion. The so-called Harvard style is used, and it has several advantages*.

Reference *citations in the main text* should be made as in the following examples: "It was shown by Curtis and Osborne (1966) that ..." or "It has been shown elsewhere (Paine, 1966) that ..." or "it may beshown (e.g. see Knuth, 1973a) that ...", but not "as described in (Hoare, 1973), the result is...". Note that the citation is composed of the author's last name and the year of publication. Where this is ambiguous, different works by the same author in the same year are distinguished by adding a single lower case letter to the year. Only one level of parenthesis is used. The author's name appears outside the parenthesis if the reference is direct (i.e. if the author's name has a syntactically valid reason to appear at that point in the sentence), or inside if the reference is indirect (i.e. if the name isn't a correct component of the sentence). In the latter case, the name and year are separated by a comma. If there are two authors, both names are used as part of the reference. If there are three or more authors, then all authors' family names should appear in the initial citation, but subsequent citations may be abbreviated by replacing the second and later author names by the phrase "*et al.*". Thus, for example, a second citation of the book "Newey, Stanton and Wolfendale (1978)" may be made as "Newey *et al.* (1978)". In this style, the citations are not intrusive; they are often sufficient for the reader to recognise the work without further effort, and they may be added or deleted easily during drafting without disturbing any pre-arranged number scheme.

References to *unpublished works* or *private communications* should be avoided as far as possible. If these are to appear, they should appear within the main text as, for example, "Lone and Ryder (to appear)" or "F.G. Smith (private communication)". They should not appear in the final reference list unless publication has already been arranged.

A section at the end of the report headed "References" should quote full details for all references cited in the text (and no others). These must be sufficient to allow an ordinary reader to locate the reference without undue difficulty, and they should be arranged so that the textual citations can be located readily. I.e., the reference list should be ordered alphabetically by author name and then year. To this end, the publication year (together with any alphabetic letter suffix) is listed immediately after the author's name and before the title of the work and the remaining details. The following items illustrate the style of a reference list:

BINGHAM, J.A.C (1967): An Improvement to Iterative Methods of Polynomial Factorisation, *Commun. ACM*, **10**, pp. 57 - 60.

CLOCKSIN, W.F. and MELLISH, C.S. (1981): *Programming in Prolog*, Springer-Verlag, Berlin.

* The material relating to referencing in the remainderof this section has mostly been copied directly from the Australian Computer Journal's Information for Authors section, which appears in the journal occasionally, is not attributed to any author, has no fixed location, and is thus totally impossible to reference according to its own guidelines.

FERNANDEZ, J.I. (1984): Protocol Translation for Packet Network Interconnection, *Aust. Comput. J.*, **16**, pp. 14 - 21.

KNUTH, D.E. (1973a): *The Art of Computer Programming, Vol 1, Fundamental Algorithms,* Second edition, p. 325, Addison-Wesley, Menlo Park, California.

NEWEY, M.C., STANTON, R.B., and WOLFENDALE, G.L. (eds.) (1978): *Programming Language Systems,* Australian National University Press, Canberra.

PAINE, R.M. (1966): Preparation for Optical Character Recognition, *Comput. J.*, **9,** pp. 221 - 229.

SNYDER, L. (1982): Introduction to the Configurable, Highly Parallel Computer, *Comput. J.*, **15**, pp. 47 - 56.

Authors are responsible for ensuring the accuracy of all details of all references quoted. Guidelines to be observed during the preparation of the reference list are: Author names are capitalised (for prominence) with initials following the family name; the year of publication is enclosed in parentheses and followed by a colon; for references to *books* and *monographs*, the colon is followed by the title (*italicised*), a page reference (if appropriate), the name of the publisher and the place of publication; for references to *journal articles*, the colon is followed by the full title of the article (without quotation marks), the name, or a standard for the name, of the journal (*italicised*), the volume number (**emboldened**), the issue within the volume, and the page number range (introduced by the symbol "p." for a single-page range, or the symbol "pp." otherwise).

Every effort should be made to obtain all of this information, as it is the only way a reader can be sure of locating the article to which you are referring. If the author of an article is not listed, or citing the author's name would not assist in locating the article (this might occur in a set of collected articles published as a book, for example), the editor/s (followed in the bibliography by "ed." or "eds." should take the place of the author/s. Occasionally, articles are not attributed to any one person or identifiable group of people, and the name of the issuing organisation may, as a last resort, have to be used.

Titles consisting of a single word are never abbreviated, and a leading *The* is usually omitted. For longer titles, the following abbreviations should be used: *ACM* (Association for Computing Machinery), *Abstr.* (Abstracts), *Appl.* (Applied, Applications), *Bull.* (Bulletin), *Commun.* (Communication[s]), *Comput.* (Computer[s], Computing), *Des.* (Design), *Eng.* (Engineering), *IEEE* (Institute of Electrical and Electronic Engineers), *Inf.* (Information), *J.* (Journal), *Math.* (Mathematics), *N.Z.* (New Zealand), *Proc.* (Proceedings), *Program.* (Programs, Programming), *Rev.* (Review), *Sci.* (Science(s), Scientist), *Softw.* (Software), *Surv.* (Surveys), *Syst.* (Systems), *Trans.* (Transactions). Thus sample abbreviations are *Commun. ACM*, *Datamation*, *IEEE Trans. Softw. Eng.*, and *Mass. Inf. Sci. Rep.*

Finerman, A. (1981) has published a more extensive summary of accepted abbreviations in accordance with ISO 833, *Documentation - International list of periodical title word abbreviations*.

When should I refer to The Literature?

There are two considerations to take into account when you are deciding whether or not to include a reference to someone else's work or writing.

First, since you're writing academic material, you need to be meticulous about referencing any material that you didn't develop yourself. There are lots of reasons for this. You want to give credit where it's due. You want your readers to be able to follow up an intriguing reference. You want your readers to be able to check that your interpretation is consistent with what the original authors actually said. You want to prevent anyone (like your assessor) from being able to say that you have claimed credit for other people's ideas.

Secondly, you need to tailor your references to the knowledge level of your target reader. You should include references to anything that falls outside their likely information base. This can result in some interesting decisions. For example, if you are including information that is basic in one area (such as nuclear physics) but you are writing for a readership from a different area (such as Computer Science), it may be appropriate to include lower-level information than a Nuclear Physics student would include in a thesis, and references to sources that are at a comparatively low level such as text books.

More commonly, you are writing for a readership which can be assumed to have a general education in your subject, though not necessarily in your specific research field. In such circumstances, it's a good idea to reference any material that falls outside the scope of a degree course with a specialisation in your general area. For example, in a Computer Science thesis concerning the development of a new interface paradigm, you would be able to assume that your readers will have a general knowledge of computers and software, and some specialised knowledge of the principles of HCI. However, you'd probably want to include specific references to important papers about HCI, particularly where your approach has been influenced by principles included in those papers.

You should *always* use references …

when you're using someone else's results or data

when you're using someone else's ideas or opinions

when you're quoting someone else directly.

This applies even for short quotations. Someone told me once that the odds that two people will use an identical sequence of words to express the same idea becomes vanishingly small if the number of words exceeds six. I don't know where this information came from, but it has the ring of truth about it. So make a definite rule: if you quote more than six words from another document, include a reference. *This includes information copied from the Web.*

This six-word rule doesn't exempt you from referencing shorter quotes. For example, if someone has invented a shorter phrase that captures the essence of a subtle situation, you would probably want to acknowledge their contribution to the subject, and your writing. Of course, when you're introducing new technical terms this will even extend to individual words.

when you've copied or *included a modified copy of* someone else's diagram or graph

when a reader in your target group, who may wish to investigate something further won't know where to look.

References

FINERMAN, A. (editor-in-chief) (1988): *Comput. Rev.*, **29**, pp. 609 - 621, 1988