

Handling Text and Type on a Computer

Aidan Kane
September 2007

Department of Economics
NUI Galway
Ireland
www.economics.nuigalway.ie

Abstract

These notes deal with a range of issues relating to the intelligent handling of text and type on computers. The general rules are to aim for readability and consistency in presentation, considerations which are informed by the art of typography.

Keywords research, writing, typography

JEL Classification Nos. A20, C88

Contents

1	Introduction	1
2	Keyboard characters	1
3	Fonts	2
3.1	Varieties of fonts	2
3.2	Serif and sans serif fonts	3
3.3	Proportional and monospaced fonts	3
3.4	Styles: bold, italics, upper and lower case	5
3.5	Font sizes	7
4	Spacing and alignment of text	7
4.1	The how and why of spacing	7
5	Predefined styles, logical vs. visual mark-up	10
6	Consistent text formatting	11
7	Conclusion	12

1 Introduction

These notes deal with a range of issues relating to the intelligent use of the text formatting capacities of word processing software. Text which is simply and consistently formatted communicates effectively. Text which is over-formatted in an attempt to ‘look well’, usually fails to communicate, by obscuring the substance in unnecessary decoration. A jumble of font sizes, styles and types is useful only to typesetters of *The Face* and authors of ransom notes. The following sections explain some basic concepts and conventions in typography which can help format text to acceptable standards of consistency and clarity.

2 Keyboard characters

Most keyboard characters will be familiar to you, but the following are sometimes confused; 1 (one) and l (the lowercase letter) and 0 (zero) and O (the capital letter). When a figure like £210,000 is typeset, the ugliness of the result should be obvious, at least in contrast to the correct version: £210,000. Such errors are irritating at best—in spreadsheet programmes they will lead to numerous puzzling problems.

Slightly more subtle issues arise in the case of characters used to represent various types of dashes. There are four typographically distinct dashes:

the hyphen	-	used between parts of the same word e.g., anti-smoking pests
the ‘en’ dash	–	separates dates and page numbers e.g., 1978–79 or pp. 30–35
the ‘em’ dash	—	separates clauses in a sentence—like this.
the minus sign	-	is what you see when you confuse your dashes: -10%

Most word processors allow the first three types of dashes to be made through various key combinations. (Packages such as L^AT_EX also distinguish the minus sign, as in the example above.) For both Mac and IBM-compatible PC versions of MICROSOFT WORD 6.0 (and later) under the **Insert** menu you can select **Symbol...** and choose ‘en’ or ‘em’ dashes, as well as a number of other special characters not immediately accessible from the keyboard. The

minus sign can be inserted using the hyphen key, as normal. No spaces should surround a hyphen, i.e., semi-circle, not semi - circle. While conventions differ with the two longer dashes, it is a reasonable rule not to surround them with spaces either—as in the examples above, and as in this one.

Quotation marks to the left and the right of a piece of text are typographically distinct; ‘like this’. These are known as ‘smart quotes’ and your word processor usually has an option to automatically insert them when you press the quotation mark key. The correct right quotation mark serves also as the apostrophe. Note that no spacing between the quotation marks and the text which they enclose is used. Double quotation marks in the correct form appear in the form of sixty-six and ninety-nine; “like this”.

3 Fonts

3.1 Varieties of fonts

Whichever character is being typeset, a very basic design decision is to choose from the wide variety of fonts, i.e., type designs. Some of the most popular fonts are:

This is an example of Palatino

This is an example of Helvetica

`This is an example of Courier`

This is an example of New Century Schoolbook

This is an example of Times Roman

This is an example of Computer Modern

The choice of particular fonts can drastically affect the readability of text. As a rule, avoid using more than three different fonts in a document (not counting special fonts for mathematical typesetting). What principles might inform the choice of fonts? At a basic level, it can be useful to understand the difference between serif and sans serif fonts, and the difference between proportional and monospaced fonts.

3.2 Serif and sans serif fonts

Serifs are the small strokes found at the extremities of individual letters. Serif fonts have these strokes, sans serif fonts do not, as should be clear from the following enlarged examples:

T s f A Times Roman font (a serif font)
T s f A Helvetica font (a sans serif font)

T s f A Computer Modern font (serif version)
T s f A Computer Modern font (sans serif version)

Sans serif fonts looker ‘cleaner’ than serif fonts, and are an excellent choice for section headings and for text/numbers which appears in graphs and tables. However, one theory says that the extra detail in serif fonts gives the eye more information with which to distinguish the shapes of letters and words, thus making blocks of text in serif fonts easier to read. A good general rule is to a serif font such as Times and Palatino for your main text and reserve sans serif fonts for headings.

3.3 Proportional and monospaced fonts

A second useful distinction is between proportional and monospaced fonts. The characters in proportional fonts vary in width, e.g., ‘m’ is about twice as wide as ‘n’, and ‘n’ in turn take up more space on the line than ‘i’. Monospaced fonts, on the other hand, devote equal horizontal spaces to each character. The following example illustrates this difference:

mmmmm Proportional (Times Roman)

nnnnn

iiii

mmmmm Monospaced (Courier)

nnnnn

iiii

Blocks of text in proportional fonts, it is sometimes argued, present the eye with more visual information, and thus are easier to read, in that they allow us to recognise more rapidly the distinctive shapes of individual words. This would suggest that the main text should always be typeset in a proportional font, and that fonts such as Courier should be avoided for this purpose. However, monospaced fonts, precisely because they give equal space to each character, are useful in other contexts. They are generally used when precision in reproducing individual key strokes from written text may be required, which is why email and web addresses as well as listings of computer code are often typeset in a monospaced font e.g., <http://www.nuigalway.ie/ecn/>.

Monospaced fonts are sometimes a good choice for typesetting large tables of data. For example, numbers which do not align at the decimal point neatly below each other in a font like Times Roman may well do so in Courier. This is often the case with data sets or text retrieved from the internet or received on disk. Data sets are often in a format with a fixed number of decimal places and a fixed number of spaces separating each column. If the original data appears to be out of order, try changing the font to Courier and reducing its size to 10 points or 9 points. The example below illustrates this, showing two versions of the same small table. In the first case, the bottom row is misaligned, whereas the second version, this problem is corrected simply by changing the font, and not by adding extra spaces.

2.164 7.413 Proportional (Times Roman)
 3.456 6.523
 5.89 7.514

2.164 7.413 Monospaced (Courier)
 3.456 6.523
 5.89 7.514

Apple Mac users should note that they should avoid using fonts which are named after cities e.g., Geneva, Chicago, Monaco etc., for printing (as opposed to viewing the files on screen). These fonts do not print as fully formed smooth letters on some laser printers, in addition to being both sans serif and monospaced, which may be undesirable for reasons alluded to above.

3.4 Styles: bold, italics, upper and lower case

Any particular font may be typeset in a number of styles or faces, in either upper or lowercase letters. The next example illustrates the options for the Computer Modern font used throughout this document

abc ABC normal
abcdef ABC bold
abcdef ABC italics
abcdef ABC slanted

Note that in general, italicising a font does more than simply slant the characters forward; 'a' and 'f' differ in shape from 'a' and 'f', as in the example above. One other variation is the style of font known as small caps. Small caps are distinct from upper case letters:

Department of Economics	Upper and lower case letters
DEPARTMENT OF ECONOMICS	All upper case letters
DEPARTMENT OF ECONOMICS	All small caps

Considerations of readability and consistency also inform the use of these various styles of fonts. Using upper and lower case letters according to normal capitalisation conventions provides the most readable text. IN CONTRAST TEXT WHICH IS TYPESET ENTIRELY IN UPPER CASE PRESENTS THE EYE WITH A SERIES OF RECTANGULAR SHAPES TO READ, WHICH RAPIDLY BECOMES TIRING ON THE EYES. Do not set text entirely in upper case. It can be acceptable in headings and titles, but is probably best avoided. Small capitals should be used sparingly. They are used in this document to to give emphasis to the names of specific software packages, e.g., MICROSFOT EXCEL.

Text which is typeset in **bold** stands out without being offensive to the eye, and is a good choice for formatting main section headings. Italics are often used to distinguish foreign language phrases, such as *a priori*, *ceteris paribus*, and the like, but they are most useful for emphasising *individual words* or phrases. Using italics in this way can often eliminate the need to enclose words in quotation marks and is a much preferred alternative to using uppercase letters or more elaborate typographical tricks.

Never underline anything in your text. The underlining feature of word processors is essentially a nuisance. The reason is that underlining a word cuts through the lower part of characters such as ‘g’, ‘f’, ‘p’ and ‘q’ (i.e., characters with *descenders*), thus breaking up the shape of the word on the page unnecessarily and impairing the readability of the text. One does see underlining used, but almost never in professionally produced documents. In the days of manual typewriters, and before, when an author marked a manuscript by underlining words, that was understood by the typesetter to mean ‘emphasise these words’, which would be achieved by italicising them, or by setting them in bold type, options not then available to the author. Underlining is particularly pointless in headings, where the size of the text alone should provide sufficient visual clues as to the logical status of the text. Underlining is a convention on the web, signalling a link to another document, which is an unfortunate feature in design terms. Another exception to this prohibition is where the underlining feature of your software is an intelligent one, as in the L^AT_EXpackage. In general, though, lose the underlining.

3.5 Font sizes

A font is large or small relative to the length of the standard line of text on the page, and the logical status of the text. Thus in short newspaper columns, text larger than 7 or 8 points would be inappropriate, leading to too many words being hyphenated at the end of each line, and so impairing readability. Most word processors default to 12 point Times Roman font which is judged by many to be too large relative to the standard line length on an A4 page. For most body text and for many fonts, a size of 10 or 11 points would be more visually appropriate, especially combined with one and a half inter-line spacing. If such sizes look well when printed from your computer, but are too small to type comfortably, you can try to increase the magnification at which you view your document on-screen. Alternatively, enter your text in a large size for ease of viewing (say 14 point) and reformat it just before printing. As indicated earlier, the relative importance of section headings can be signalled by varying the size of text, but in no case should you use a point size higher than 14. Text in footnotes and endnotes is often set automatically in a lower point size than the main body text, but this is not essential.

4 Spacing and alignment of text

Spacing and aligning text properly can increase readability a great deal. The following general guidelines tells you how. The discussion which accompanies them tells you why.

4.1 The how and why of spacing

‘Justify’ the text of the main body of your paper. That is, align it so that it is flush with both the left and right margins of the page, as in most of the paragraphs in this document. When your word processor justifies text, it adjusts the spacing between words on each line to fit the text exactly on that line, hyphenating words according to a set of in-built rules when necessary. Justifying text can produces lines in which words are ‘strung out’ between ugly blocks of white space if the point size of the text is too large relative to the line length. This suggests looking at fonts of 10–11 points

size, rather than only at 12 points, for the main text. Note also that because monospaced fonts have characters of equal width, the word processor has much more difficulty in justifying text set in such fonts, which is another reason to avoid them for most purposes.

Use standard spacing for punctuation. That is, one space only between words, and no space before, and just one space after punctuation marks such as commas, exclamation marks, question marks, colons, semi-colons, full stops (periods) and so on e.g, this!, and not this ! and not this ? but this? and this, not this , and not this . but this. Note in particular that one space only should follow a full-stop, not the two spaces which typists trained on manual machines often use. Most software will automatically insert slightly more space after a full stop, which is enough to distinguish individual sentences.

Most spacing problems arise when this rule is not followed, as the word processor attempts to add too much extra space when justifying text, resulting in words strung out on a line. Near the end of your draft, use the 'search text' facility on your word processor to remove extra spaces. If in doubt about a particular point of punctuation, leave the space out.

Use standard margins. That is, the margins set by default in your word processor are generally appropriate. The logic of setting margins wider than might be possible given the size of the paper is to give a line length which is not too long to read efficiently, to allow for binding, to allow for marginal notes to be made, and to allow photocopies to be made easily. One suspects some texts are published with awkward page sizes and margins to inhibit photocopying.

Use one and a half line spacing in the text for maximum readability and to allow space for corrections and notes as appropriate to be marked by hand. Single spacing should be used for items in numbered lists, bulleted lists and bibliographies.

Leave a blank line (press the 'return' key) between each paragraph. This suffices to distinguish each paragraph and there is then no need to indent the first line, which is unduly fussy for most purposes. Exceptions again are for items in numbered lists, bullet lists and bibliographies.

Do not use the space bar for spacing blocks of text. This rule can save you a lot of trouble. The only function of the space bar is to separate words from each other. It should not be used to try to align text from the

margins. For this purpose adjust the margins directly (check the ‘ruler’ which your software probably displays for the symbols which mark the margins), or use the alignment function on your word processor (the options being left aligned, right aligned or justified text).

Both right and left margins of block quotations are indented, and the left margins of items in bulleted or numbered lists should be indented (see examples in this document). Using the space bar and/or lots of tabs to adjust margins does not work partly because characters in proportional fonts vary in width, so that it is impossible to predict how many spaces/tabs one needs to force a particular block of text to the desired position. Above all, if and when you change that text, the alignment generally turns into a mess. This will not happen if you use the space bar to once only between words. A case which usually causes authors problems is trying to format . . .

Hanging indents in which the first line of a paragraph is closer to the left margin than are all subsequent lines. Which is clearly the case in this paragraph. Bibliographical entries are sometimes typeset like this, in order to make each item distinct on the page. A set of definitions could also be typeset in paragraphs with hanging indents.

A paragraph which hangs by one tab stop on the first line can be used to make numbered lists and lists of bullet points by inserting one tab stop between the numeral or bullet symbol and first word of the first line of the text. If one then changes the text in any item on the list, it will continue to align correctly on the page. If the text had been forced from the margin by inserting multiple tab stops or spaces across lines, changing the text would undoubtedly mean further messing with tab stops and spaces.

Use tab stops carefully. Tab stops are properly used mainly to separate distinct items at specified intervals on a line, most obviously in tables of data. Each column entry in a row should be separated by one tab stop. If initial results do not align properly, adjust the distances between each tab stop (usually from the document’s ‘ruler’), rather than adding extra tab stops. This helps to make the text robust to any editing changes and ensures that the document can be easily translated between different formats e.g., from a word processed table to a spreadsheet file, or to a data file for a statistical package, or to a table for a web page.

Most word processors allow you to view your document showing tab stops, spaces generated by the space bar and the ends of paragraphs as non-printing

symbols on the screen (arrows for tab stops, grey dots for spaces, and ¶ for paragraphs). Viewing such ‘invisibles’ will usually reveal the sources of misalignments as overuse of the space bar and tab key.

5 Predefined styles, logical vs. visual mark-up

Many users of word processors are unaware of these distinctions, which leads to needless frustrations with the technology. A way out of these problems is to:

Use the style facility on your word processor. Spacing and indeed most formatting issues for your text can be handled very simply by taking advantage of the pre-defined formats on most word processors. These are generally termed ‘styles’ which one can apply to selected paragraphs of text. For example, the style for ‘Heading 1’, the most important section heading, might involve 14 point bold Times Roman text with no indents, left aligned, single inter-line spacing and 14 points of white space after the text. The author need not be concerned with specifying such a list of typographical settings. All he/she needs to do is to select the relevant text in the document and ‘tag’ it via a menu as ‘Heading 1’. If you do this to all your main section headings, then if for some reason you wish to change the style definition of ‘Heading 1’—say if you wish it to be set in Helvetica font—then all the section headings in your document will be changed at the same time. A related feature of some such pre-defined styles allows you to number headings automatically and generate tables of contents with correct page numbers by choosing one item on a menu. Most word processors also have collections of such styles appropriate for different document types (e.g., CVs, letters, theses) known as document templates. Learning about these features is well worth the initial effort. Details can be usually be found in the on-line help facility and/or documentation of the software in question.

The fundamental and general idea here is that one has used ‘logical mark-up’, rather than ‘visual mark-up’ only. This means that when text is tagged according to whether it is a section, sub-section or body text etc., the document then embodies information about its own logical structure which can be used in other contexts, such as generating structured tables of contents. A document which is structured logically in this specific sense of its com-

ponents being tagged can also be easily translated between other document formats. For example, the job of translating this document from L^AT_EX to its web version—a format known as HTML (HyperText Mark-Up Language)—was a matter of a few minutes. Similarly, a document tagged appropriately in a presentation/slide software package such as MICROSOFT POWERPOINT will retain its logical structure when opened from a word processor such as MICROSOFT WORD—the title of individual slides are read as main sections, bullet points on slides are read as sub-sections, and so on. The utility of the concept of logical mark-up becomes rapidly apparent as you use a greater variety of software tools for different purposes.¹

6 Consistent text formatting

Watch your apostrophes. The apostrophe is used to denote possession (the chancellor’s statement), or to stand for missing letters (won’t, don’t). But note that ‘it’s’ is short for ‘it is’ and there is no apostrophe in the word ‘its’ when it’s used in the possessive sense as in: ‘the world and its wife gets this rule wrong’.

Avoid excessive capitalisation. Start a sentence with a capital letter, and also proper nouns, e.g., the names of countries, persons etc. Do not capitalise words merely because they seem important. Note the limited capitalisation in section headings in this document, and in the table of contents.

Use minimal punctuation with acronyms e.g. EU rather than E.U. Also, on the first occasion an acronym is used, it should be spelled out in brackets immediately afterwards e.g., “the IMF (International Monetary Fund) has implemented a number of structural adjustment programmes.” Don’t needlessly generate acronyms of your own, and avoid any unnecessary ones: IT is unnecessary, even OTT; information technology will do just fine.

Question marks should be used when asking or quoting a question asked directly, and not when reporting one. The following sentence reports a question. Keynes asked whether the economy would reach equilibrium without

¹An excellent example of the utility of logical mark-up is that web pages which are correctly structured according to HTML can be sensibly interpreted by software which reads aloud them via a synthesised voice to people whose vision is impaired. What might be thought of as an intrinsically visual document then does not have to be seen; it can be heard.

government intervention. No question mark is necessary. Also, the use of the word ‘whether’ is more grammatically correct than using the word ‘if’ in its place in such sentences.

Adopt consistent usage. It would matter little whether you write U.S.A., USA, US or America , but stick with one of these (U S of A would be a bit much).

A set of decisions on such conventions comprise a ‘house style’. Other examples would be:

- for referring to decades, use 1980s, 1970s rather than 1980’s, 1970s, or the eighties, the seventies.
- for percentages, use the percentage sign, rather than words i.e., 12%, rather than twelve per cent or 12 per cent.
- the following punctuation i.e., and e.g., rather than i.e. and e.g. and definitely not ie and eg.

As you draft, keep a running list of any such style decisions you make, so you can be sure you are consistent throughout.

Adopt standard UK English spelling, rather than American English, except when quoting directly or giving the title of a publication etc.

7 Conclusion

These notes may help to clarify some of the basic issues in handling text and type on a computer. If not, consult other sources, including on-line help facilities in your software, and let the ultimate test of the presentation of your work be its readability, rather than a set of rigid rules.