First principles of typographic design for document production

Richard Southall

1 Introduction

Leslie Lamport and I taught a two-day course on 'First principles of typographic design for document production' as a preliminary to the TUG meeting at Stanford University, August 13–14, 1984.

What follows is an expansion, somewhat revised and restructured, of my lecture notes.

Objectives

I did not feel that it was possible in two days to teach anything useful about typographic design on a 'how to' basis (even if the problems of interpreting conventional typographic design practice in terms of TEX macros had been solved). Nor would it have been helpful to give preceptive solutions to a limited set of design problems. On the old analogy, I did not wish to give away fish to the course members, and I did not have time to teach them how to catch fish for themselves. In such circumstances, the best thing to do is to try to outline some productive ways of thinking about water and fishing-tackle, so that basic errors of approach to the problems of fish-catching can be avoided in the future. Thus I tried to work out, and justify from first principles, a useful way of thinking about the problems of document design; and to illustrate this way of thinking by looking at the ways in which typographic designers working in earlier technologies had tackled similar kinds of problem.

Subject area

The subject matter of my part of the course was written language.

Written language has two levels of structure, which may be called microstructure and macrostructure (Waller, 1980a). The microstructure of written language has to do with details of the sequence and arrangement of the characters that make up a written text. It is governed by the rules contained in a 'system of writing': the set of rules for writing a particular language with a particular script in a particular technological environment. (I owe this useful concept to John Mountford.) For most systems of writing, and certainly for those in which the written text is produced by a mechanical writing-system, these rules are reasonably well defined, and descriptions of them are readily accessible (Chaundy et al., 1957; 'Chicago manual of style', 1982; Dowding, 1966; Swanson, 1979; Walker, 1979); though they differ to a surprising extent between systems of writing, even in the same technological environment (Desarmenien, 1984; Walker, 1983).

The macrostructure of written language has to do with its division into semantic objects (which we can usually recognize, even if we can't make very exact definitions of them — things like chapters, subsections, paragraphs and list items) and their graphic embodiment in a document. The rules governing the macrostructure of written language are much less well defined than those governing

its microstructure. Many typographic designers would deny that there were any such rules at all (though Twyman, 1981, formulates 'what many typographers would consider one of the few fundamental "rules" of typography'), because there are so many equally effective ways of designing a document. But if there are no rules (or almost none) that designers would agree on, there are some principles for effective design that they (or most of them) would agree with; and it was these principles that were the subject area of my part of the course.

Our discussions in the course were based on two postulates:

> Documents have a conceptual structure
> Graphic structures can be made that reflect conceptual structures

one axiom:

> The graphic structure of a document should reflect its conceptual structure

and one desideratum:

> The graphic structure of a document should be such that the document is as easy as possible to use.

The problem of document design for computer-based systems

In the past, the design of documents was done *a posteriori*. Books were written before they were designed: the conceptual structure of the author's thoughts was already in place, embodied in some graphic form or other, and the designer's task was to render or re-render that embodiment in a semantically effective and technically practicable way. (See Hewson, 1983, for a detailed study of the evolution of the graphic embodiment of Wittgenstein's 'Tractatus Logico-Philosophicus' through successive renderings — in manuscript, typescript and printed form — of the work.)

In the design of documents for computer-based document production systems, the problem is the other way round. The document has to be designed *a priori*, before the author's thoughts are present at all. The document designer's task is to devise efficient graphic embodiments for conceptual structures that are suitable to fit any thoughts that any author using the system might have.

The document designer cannot tell (let alone dictate) what content an author will put into each of the conceptual structures that the document design provides for, or in what order the structures will be used. It is not too hard to provide a graphic embodiment for each structure, that will behave reasonably if it is not used in what its designer would consider to be an unreasonable way; but what is unreasonable to a document designer may not be at all so to a mathematician or a philosopher.

2 Typographic structures

Models of text and models of documents

There is a simple 'bottom-up' model of written text, that considers it as a sequence of characters and spaces.

Sequences of characters make up words

words	sentences
sentences	paragraphs
paragraphs	sections
sections	chapters
chapters	books (articles, reports, documents)

This model does not work very well, either as a way of representing real text or real documents. At the lowest level, it ignores things like changes in character style. At a slightly higher level, it doesn't deal at all tidily with quite important questions, like 'What is a sentence?' (or, indeed, 'What is a word?'). At a higher level still, it ignores many of the important things that aren't chapters (in particular, lists and tables) that go to make up documents.

As a means of understanding the nature of written text, this model is good for analysing the functions of non-alphabetic characters in words and straightforward sentences; but otherwise it is too simple to be useful with present-day technologies of mechanical writing. It was well suited to the train-printer and daisy-wheel era of document production.

A 'top-down' approach to the modelling of documents begins by looking for recognizable graphic objects, and the semantic objects of which they are embodiments, in a document.

This approach has been pioneered in the British Library research project on the graphic translatability of text, with work by P.E.Norrish in the Department of Typography & Graphic Communication at the University of Reading. This work has centred on the study of public information documents, which are usually much more complex in both graphic and semantic terms than the technical reports that designers for computer-based systems are most often concerned with. Norrish and her collaborators have made very detailed analyses of such documents, looking at the 'access structures' of headings and paragraph markers that allow users to find their way around a document, as well as the objects that occur in it.

This work is relevant to the problems of designing computer-produced documents in two ways. First, it points up the considerable complexity, in semantic as well as graphic terms, of real documents. It is all too easy to think that chapters, sections and subsections are the only kinds of object that need to be considered in designing the text of a document. Second, it draws attention to the large number of alternative graphic forms in which a particular semantic object can be embodied. Norrish has so far identified no fewer than forty-seven different types of 'graphic list' (semantic list structures in which the list items are distinguished by graphic means).

Section structures

However inadequate the chapter/section/subsection model may be as a way of representing the structure of real documents, it is very often used to construct the body matter of technical reports. If we look at some of the semantic properties of constructions of this kind, we can get an idea of the properties their graphic embodiments ought to have.

Three kinds of relationship exist between the elements of a chapter/section/subsection construction. There are relationships of hierarchy: chapters are higher-level elements than sections, which are higher-level elements than subsections, and so on. There are relationships of containment: higher-level elements contain lower-level elements. And there are relationships of sequence: objects at each level follow each other in the construction.

At all the levels of such a construction, the logical structure of the elements is very similar. Each chapter, section, subsection and so on contains a heading and a body: the heading relates to the body that follows it. The body of the element may contain paragraphs of text, or lower-level elements, or text followed by lower-level elements, down to the lowest-level element, whose body (by definition) is all text. Thus there may be 'descending' sequences of headings at the beginning of an element.

By our axiom, the graphic embodiment of a chapter/section/subsection construction should reflect its semantic structure. Since the visual appearance of the paragraphs of text within an element does not usually alter with the position of the element in the hierarchy (except perhaps at the very lowest levels), it is the task of the heading of each element to show whereabouts in the structure its element is located.

Thus, the hierarchical relationships between elements should be clearly expressed in terms of the graphic relationships of their headings; as should their relationships of containment and sequence. The means that are available to designers for realizing the graphic expression of such relationships are discussed next.

3 The typographer's tools

Typeface terminology

The terms 'typeface' and 'font', which in earlier technologies had separate and clearly-defined meanings, are now used more or less interchangeably in discussions of computer-based document production systems, to refer to fundamentally different entities. Great confusion results. The following definitions are proposed, as an aid to clear thinking:

Typeface: a set of distinctive, visually related shapes for some or all of the characters of a script, intended for mechanical reproduction

Style: a distinguishing visual characteristic of a typeface

Family of typefaces: a set of visually related typefaces with differing styles

Font: a set of renderings of some or all of the character shapes of a typeface, intended for use in a restricted range of output image sizes in a particular reproducing system

What happens in text?

Written language contains elements which are not alphabetic or numeric characters: punctuation signs, and space. It also has features which seem to operate at a higher level than the words of the text: capitalization, changes in type style and size, and the presence of vertical and horizontal space in varying amounts. These elements and features evidently have some part to play in written language, and some (but surprisingly few) attempts have been made to identify their functions (Mountford, 1980: Walker, 1979). The following list of 'roles fulfilled by graphic and spatial features in the articulation of verbal graphic language' (whose incompleteness is acknowledged by its author) is taken from Walker's paper:

1 Differentiation

- 1.1 Emphasis
- 1.2 Distinction/particularization
- 1.3 Quotation
- 1.4 Interpolation

2 Abbreviation

3 Introduction

4 Omission

5 Separation and connection

6 Presentation of numbers

The most significant of these functions seems to be that of differentiation/emphasis.

Graphic conventions and graphic capability

Any mechanical writing-system makes available to its users a certain graphic capability. This can be expressed in terms of the number of characters, typefaces and type sizes, and the facilities for defining amounts of horizontal and vertical space, that the system offers (Southall, 1982). Until very recently, the graphic capability of the writing-systems that were available to most computer users was extremely limited: a single 96-character font of one single-width typeface, with a single increment of horizontal space the same as the characters' width, and a single larger increment of vertical space.

Except for the lack of a 'half-line' increment of vertical space, this capability is roughly the same as that of an ordinary typewriter. Thus when computer-based document production systems came to be designed, there were a number of ready-made graphic conventions for expressing semantic function already available in the rules for the layout of typewritten documents. Almost all of these conventions could be taken over unchanged into the layout of computer-produced

documents, and there was no need for system designers to think explicitly about the semantic functions that were being expressed by particular graphic configurations.

Laser page-printers, and the first versions of TEX and Metafont, made writing-systems with enormously increased graphic capability available to part of the computing community. The designers of document production systems did not always understand that graphic conventions derived from typewriting practice were not necessarily appropriate for systems with a graphic capability as good as, or better than, that of conventional typesetting.

The typographer's task

Looking at written language as being made up of semantically functional graphic elements as well as the words and phrases of the text, and bearing our axiom in mind, it becomes easy to define the typographic designer's task. This is to devise effective graphic means of expressing the different semantic functions that are required to embody in a document the conceptual structure of its author's work.

The way we are accustomed to reading continuous text leads us to recognize a paragraph as an area of text within which the space between successive rows of characters does not vary. Because of this, the graphic means that are available to the designer for expressing the functions of emphasis and differentiation within a paragraph are limited to those that do not change the space between rows of characters: capitalization, and changes in type style without changes in type size.

Outside the limitations of the paragraph, additional graphic means of expressing semantic function are available. The most powerful and flexible of these, and the least well understood by untrained designers, is the use of space. Others are changes of type style, and changes of type size.

4 Making text readable

Legibility research is an old subject (Tinker, 1963) and — as far as investigations of the typographic requirements for readable text are concerned — more or less a dead one (Spencer, 1969; Zachrisson, 1965). Current research in reading focusses much more on the perceptual and mental processes involved (Pirozzolo and Wittrock, 1981; Tzeng and Singer, 1981). Perhaps it is what computer scientists see as the excessive antiquity of the research on the layout of readable text that has led it to be so consistently ignored by the designers of computer-based document production systems. Its findings, though, are none the less valid for being old, or for being ignored, and they are as follows:

> Lines of text should not be longer than 10–12 words (60–72 characters)
>
> The appearing space between words in a line of text should be substantially less than the appearing space between successive lines of the text

The appearing space between words should not vary appreciably from line to line of the text.

These have been the canons of good typographic design for text for a very long time (Morison, 1951).

5 Designing headings

The functions of headings were discussed in Section 2. Headings mark the elements they belong to, and they show where those elements belong in the logical structure of the document as a whole. The latter function almost always means showing two things about an element: its place in a hierarchy and its place in a sequence.

The graphic hierarchy of the headings in a document should express the conceptual hierarchy of the elements the headings belong to. Thus the headings of higher-order elements should be graphically more prominent than the headings of lower-order elements. The connection between a heading and its element should always be explicit, so that in a sequence of 'descending' headings it is clear which element a particular heading belongs to.

In principle, the designer has three means for expressing hierarchy and connection in graphic terms: changes in type size, changes in type style, and the use of space. Where sequence is expressed explicitly, it is always by some sort of numbering system (Waller, 1980b).

In practice, the availability of graphic means to the designer of a document depends on the graphic capability of the writing-system the design is being made for. In a system with limited capability, the graphic means for expressing hierarchy may be exhausted before the bottom of the hierarchy has been reached. In such a case, the designer may choose to use a more elaborate numbering system for the headings in the document, so that the numbering expresses hierarchy as well as sequence.

This sort of circumstance, in which the content (in terms of characters and spaces) of a text is determined by the graphic capability of the system with which the text is rendered, is an instance of a major class of problems that have to be faced and solved by a realistic methodology of generalized document design.

6 Laying out the page

Margins

While our mechanisms of reading have not changed in the last few years, so that the rules for the layout of readable text can be carried over unaltered from traditional practice, the same is not necessarily true of the way we use documents.

The classical canons for the size of margins that are summarized by Tschichold (1965) apply to reading situations which were usually very different from the ones in which technical reports are used. Classically, books are codices: they open to a pair of facing pages. Books intended for continuous reading at normal reading distances

are of such a size that this pair of pages forms a visual unit for the reader. In these circumstances, a symmetrical layout for the pages makes sense.

The letter-sized page which is by far the most common in technical reports is much the same size and format as the quarto page of traditional book design. Quarto formats were largely used in four kinds of book: reference books; bibles, in which the main text was broken up into verses, and the page often crowded with cross-references and notes; service books, for use in bad light by people who were very familiar with the text; and editions de luxe, where the width of the margins was often an index of the cost of the book (and hence the wealth of its owner).

Reference books and quarto bibles are meant for consultation rather than continuous reading. Service books are intended to be read at distances that are longer than normal. Editions de luxe are for admiration rather than use. There is no particular justification for taking over the convention of symmetrically laid-out pairs of pages from such books into the design of technical reports, whose letter-size pages are seen at normal reading distance (so that the visual unit is a single page) and are as often as not photocopied and put in a ring binder (which it is stretching traditional terminology rather too far to describe as a codex).

Line length

The use of the letter-sized sheet in laser printers derives from its use in the office, where conventional margins give a line length of six and a half inches or so: with the character pitch and line spacing of ordinary typewriters, a line of 65–75 characters and an easily readable page. The same margins, with the 10 point type that is the default size in many document preparation systems, give a line of 95–115 characters: far too many to be read effectively, even if the space between the lines of text is increased (Tinker, 1963). It is true that making a very long line means that many more characters fit on the page, and fewer pages are needed for a document of a given length. The consequent reductions in cost (which, as Leslie Lamport pointed out during the course, may be more apparent than real) have to be traded off against the reduced effectiveness of the document as a means of communication.

In fact, the most efficient way of filling a page with characters is to use double-column setting. Much narrower margins are tolerable in double-column than in single-column setting, so that more of the page area can be covered with type. However, the problems of designing headings, and of integrating the non-text elements of the document into the design of the page, tend to become more difficult.

Other components of the page

Text is not the only thing that appears on the pages of technical reports. There are footnotes, tables, figures and figure captions, as well as page numbers and running headlines or footlines. All these elements need to be distinguishable in visual terms: figure captions should not disguise themselves as footnotes, or headlines as part of the text. The designer needs to pay attention to unusual circumstances — What happens if a figure falls at the foot of a page? if a table whose column headings use the same style and size of type as the headline comes at the head of a page? — rather than laying out the page as if it was only to contain continuous text.

7 Conclusions

There is nothing in the foregoing that is not perfectly obvious as soon as it is pointed out (except perhaps the requirements for readable text, and they are in easily accessible research literature). Equally, I had no difficulty, with little more than a couple of numbers of TUGboat at hand, in finding many pages whose layout made them virtually impossible to read, and many instances in which the conceptual structures of documents were being concealed, rather than revealed, by their graphic embodiment.

Why is this? Why is the average computer-produced document so hard to use (and such a miserable object, in conventional graphic-design terms)? The reasons lie in a failure of communication between computer scientists and graphic designers.

This failure has two main causes. The first is that the designers of computer-based document production systems most often have no conceptual apparatus with which to think about graphic design problems. The thinking of traditionally-educated graphic designers tends to be unarticulated, visual, and concrete, and their vocabulary oriented towards particular graphic technologies; computer scientists' thinking tends to be algorithmic, symbolic, and abstract, and their vocabulary is that of mathematics. Ways of thinking about document design that computer scientists might find congenial are only beginning to be developed in the graphic design community, as a response to the problems posed by new information-handling technologies.

The second cause of failure is at the interface between graphic designers and computer-based document production systems, and has to do with the nature of the interface itself. Because graphic designers think visually, they need visual objects to think with; if the first task a designer has is to analyse the conceptual structure of a document, the second is to sketch out possible graphic embodiments for it. A skilled designer can get a long way by making drawings, but sooner or later a design has to be tested by being produced: and it is at this point, where the specifications on the drawing have to be translated into instructions to the document production system, that communication fails. The primitive concepts that are handled by computer text formatting languages, the terms the languages use to describe them, and the commands with which they are manipulated,

are all completely alien to the experience of traditionally-educated designers.

The result, in the present state of affairs, is that traditionally-educated graphic designers *cannot use* computer-based document production systems, however much they would like to do so. The most urgent priority, for TEX and systems like it, is for interfaces to be developed so that designers can bring their skills to bear in an area where — to say the least — they are badly needed.

References

Chaundy, T.W., Barrett, P.R. and Batey, C.
The printing of mathematics
London: Oxford University Press, 1957

The Chicago manual of style (13th ed)
Chicago: University of Chicago Press, 1982

Desarmenien, J.R.
How to run TEX in a French environment: hyphenation, fonts, typography (Computer science report STAN-CS-1013)
Stanford, California: Department of Computer Science, Stanford University (1984)

Dowding, G.
Finer points in the spacing and arrangement of type
London: Wace, 1966

Hewson, R.L.
The visual presentation of a philosophical text: Wittgenstein's Tractatus Logico-Philosophicus
Unpublished dissertation: Department of Typography & Graphic Communication, University of Reading, England (1983)

Morison, S.
First principles of typography
Cambridge: Cambridge University Press, 1951

Mountford, J.
Writing-system as a concept in linguistics
Information design journal, vol.1 no.4, pp.223–231 (1980)

Pirozzolo, F.J. and Wittrock, M.C. (eds)
Neuropsychological and cognitive processes in reading
New York: Academic Press, 1981

Southall, R.
The problems of forms artwork production
London: Management & Personnel Office, HM Government, 1982

Spencer, H.
The visible word
New York: Hastings House, 1969 (2nd ed)

Swanson, E.
Mathematics into type
Providence, R.I.: American Mathematical Society, 1979

Tinker, M.A.
Legibility of print
Ames, Iowa: Iowa State University Press, 1963

Tschichold, J.
'Non-arbitrary proportions of page and type area'
in Osley, A.S. (ed)
Calligraphy and palaeography: essays presented to Alfred Fairbank
London: Faber & Faber, 1965

Twyman, M.L.
Typography without words
Visible language, vol.15 no.1, pp.5–12 (1981)

Tzeng, O.J.L. and Singer, H. (eds)
Perception of print: reading research in experimental psychology
Hillsdale, N.J.: Lawrence Erlbaum, 1981

Walker, S.F.
The presentation of information about house style: a reconsideration (IET Text-processing paper 1)
Institute of Educational Technology, The Open University, Milton Keynes, England (1979)

———

Unpublished thesis: Department of Typography & Graphic Communication, University of Reading, England (1983)

Waller, R.H.W.
'Graphic aspects of complex texts: typography as macro-punctuation'
in Kolers, P.A., Wrolstad, M.E. and Bouma, H. (eds)
Processing of visible language, vol.2
New York: Plenum, 1980

———

'Notes on transforming no.4: numbering systems in text'
in Hartley, J. (ed)
The psychology of written communication: selected readings
New York: Nichols, 1980

Zachrisson, B.
Studies in the legibility of printed text
Stockholm: Almqvist & Wiksell, 1965

Bibliography
The publications cited here are chosen, as far as possible, to be good pointers into the design research literature as well as being interesting in themselves.

Green, T.R.G. and Payne, S.J.
The woolly jumper: typographic problems of concurrency in information display
Visible language, vol.16 no.4, pp.391–403 (1982)

Hartley, J.
Designing instructional text
New York: Nichols, 1978

——— (ed)
The psychology of written communication: selected readings
New York: Nichols, 1980

———
Eighty ways of improving instructional text
IEEE transactions on professional communication, vol.PC-24 no.1, pp.17–27 (1981)

Jones, J.C.
Design methods: seeds of human futures
London: Wiley-Interscience, 1970

Macdonald-Ross, M. and Smith, E.B.
Graphics in text: a bibliography
Institute of Educational Technology, The Open University, Milton Keynes, England (1977)

Reynolds, L.
Legibility studies
Journal of documentation, vol.35 no.4, pp.307–340 (1979)

Spencer, H.
Pioneers of modern typography
London: Lund Humphries, 1969

———, Reynolds, L. and Coe, B.
Spatial and typographic coding in printed bibliographical materials
Journal of documentation, vol.31 no.1, pp.59–70 (1975)

Tschichold, J.
Asymmetric typography
New York: Reinhold, 1967

Twyman, M.L.
'A schema for the study of graphic language'
in Kolers, P.A., Wrolstad, M.E. and Bouma, H. (eds)
Processing of visible language, vol.1
New York: Plenum, 1979

Waller, R.H.W.
Journal typography in transition
Journal of research communication studies, vol.3 no.4, pp.335–349 (1982)

Wright, P.
'Usability: the criterion for written information'
in Kolers, P.A., Wrolstad, M.E. and Bouma, H. (eds)
Processing of visible language, vol.2
New York: Plenum, 1980

———
"The instructions clearly state . . . . " Can't people read?
Applied ergonomics, vol.12 no.3, pp.131–141 (1981)