What Characterizes Pictures and Text?[1]

Dr Michael A R Biggs

Faculty of Art and Design

University of Hertfordshire

College Lane

Hatfield

AL10 9AB

UK

Abstract

This paper addresses an apparently trivial question: what is the difference between graphics and text? It appears to be trivial because there appears to be several alternative and simple ways of answering it. For example, 'text is made up of letters whereas graphics are not', 'one can create text using a keyboard', 'one can read text aloud', etc. However, none of these provides robust conditions to differentiate graphics from text, e.g. cases such as typewriter art and gobbledygook can be identified.

The paper approaches the problem of identifying content conditions by analysing boundary cases which lie on the margins and are difficult to classify. It considers examples that arise in the production of materials, including bitmapped text, graphics consisting of letters and words, text used as patterns or in tables, etc. It also considers examples that arise from the consumption of materials, including a comparison of the methods used for reading and interpreting text and graphics.

This paper concludes that current XML specifications, e.g. TEI guidelines, for the integration of graphics into text are primarily made on the basis of form rather than content. This is incompatible with a content-based markup scheme. Before such guidelines can be modified we must be clearer about what differentiates graphics from text in terms of content conditions rather than a technological or formal conditions.

Let me start with a simple statement of the problem. When one encodes a source in a descriptive markup language one must identify features, describe them, and place that description in the appropriate place in the code string. In typical cases of graphics this will involve inserting a reference to an external graphical file type such as a JPEG[2] into a string of text. However, being-a-JPEG is not synonymous with being-a-graphic, for example, a scanned typescript is graphical in file type but textual in content. Similarly, SVG[3] is textual in file type but graphical in content. Therefore if we are to be consistent in applying descriptive markup we should be able to differentiate between graphical content and graphical file types, textual content and textual file types, and then describe as graphics that which has graphical content. In practice this is not as easy as it sounds because there always seem to be exceptions to any description one might make of what constitutes graphical content. For example, is 'colon right-parenthesis' a smiling-face graphic[4], or is it only graphical when we use a symbol, e.g. Unicode 263A? 'Typewriter art' creates an image using typewriter characters and these may be arranged in horizontal lines. However, we distinguish between this and gobbledygook, i.e. meaningless strings of characters. We could also imagine a page in which the pattern of the text coincidentally resembled a face. Brand names such as 'Coca-Cola' are often given a particular appearance to form graphical trademarks. If we agree that all of these could be said to have graphical content then graphics cannot be defined simply by 'line type' or by being 'pictorial'. According to Mitchell 'we still do not know what exactly pictures are' (1986: 13).

On the basis that the literature about 'what is a text' (Gelb 1963, Coulmas 1990, deRose 1990, etc.) is more extensive than the literature on 'what is a graphic' (Doblin 1980, Mitchell 1986, Biggs 1995, etc.), I shall consider some descriptions of text to determine by reduction what might be a description of a graphic. The problem of describing document features has been approached reductively before, i.e. Sperberg-

McQueen, Huitfeldt & Renear (2000: 217), but little work has been undertaken on the boundary between graphics and text, and how such a boundary might be drawn for the purposes of document description, e.g. in XML[5]. Most of the commentary on text encoding assumes a textual context, and considers the problem to consist of two parts: providing an external reference and providing a textual description of the graphical content. For example, the TEI[6] guidelines on descriptive markup (Sperberg-McQueen & Burnard 2001: §22.3) include recommendations about how to markup 'figures'. It assumes textual content into which are inserted graphical file types. These external references and other source information are marked as <figure> but in so doing they fail to describe the content as opposed to the form. The question therefore remains, what is textual content and what is graphical content, and if we are to describe or markup a source, how might we determine which features to mark as which?

I will assume that we are interested in the interpretation and description of a manuscript source that may contain graphics and text, and other forms of notation common in Western Europe, e.g. European character sets such as Latin and Greek, European languages, mathematical, logical and musical notation, etc. The purpose of this restricted scope is not because the discussion is limited to such contexts but because the description of the diversity of features in other contexts would make this paper too long. I believe that the problem is transferable to non-Western contexts and syllabic rather than phonetic writing systems, etc. The purpose of differentiating features in any source document is in order to apply descriptive markup.

The Problem

Writing and drawing are everyday activities: the simple acts of making communicative marks on paper. We do not require any sophisticated equipment to write or draw: a pencil and paper are quite sufficient for both. With them we can write down a narrative, illustrate it, annotate a passage of music, and all these things we can do apparently seamlessly, just using the paper and the pencil. So at the moment of production there seems to be nothing especially remarkable or separable about the activities of writing and drawing. In particular there is no convention that we adopt for marking the beginning and the end of each notational system. But when we come to reproduce these notations using conventional technologies we find that we need different resources for different kinds of notation. The written text we convert into typography; fonts of conventionalized letterforms that we might enter via a keyboard. For musical notation we will require special typesetting or software, different from that required for text. And for images we may have to scan them from the original source document, converting them into lines of tones and colours. So at the point of reproduction some differences emerge between the tools that we require to reproduce text, non-textual conventional notations such as music, mathematics, etc., and graphics, and the boundaries between them. This description accords with Gelb's definition of writing [text] as 'a system of human intercommunication by means of conventional visible marks' (1963: 12). From this we might infer that musical notation is also a form writing but that drawing, being a system of human intercommunication by means of non-conventional visible marks, is not.

However, the fact that we might have a bitmapped image at one point, and a string of ASCII at another, tells us nothing about the content of the source. As we have seen, having a graphical file type at a certain point cannot be regarded as an indicator of graphical content at that point. For example, if I scan a printed document, does its content (as opposed to its form) become graphics? What if I create an image in SVG, does its content become text? Separating form from content accords with Coulmas's

description of writing, i.e. 'textual content', as 'linguistic content' (1990: 27). Conventional text can be read aloud, music less so, and graphics not at all. Reciting the content of an SVG file does not conjure up an image.

Finally, as readers we adopt different strategies for following the sequence of conventionalized symbols in normal text, the two-dimensional but left-right progressing notation of music on a stave, and the free-forms of graphics that our eyes may scan in whatever direction we please within the graphics boundary. This accords with Larkin and Simon's differentiation of 'sentential structure and diagrammatic structure' (1987: 66). Conventional text has a sentential structure. Musical notation has some kind of variant on that structure, and drawing uses the page surface in a completely non-sentential or diagrammatic way.

We can see that the sequence of marks across the page during writing and reading might characterize some change in mode between textual content and graphical content, and the accidents of contemporary technology might for the time being give us additional grounds to make a distinction. But what we need is some way of identifying these and other features more clearly so that (a) they can be classified and described, and (b) so that start and end points can be assigned to them in the marked-up text that is itself a linear string of code.

But even this does not fully describe the problem. Coombs, Renear and deRose (1987: basic theory) have argued that 'doing markup' consists of three stages: 'element recognition, markup selection and markup performance'. Before we can classify a feature as graphics or not, and then determine where to put start and end tags, we must first recognize and identify the feature to be encoded. I have mentioned text, musical notation, mathematical notation, and graphics. But there are perhaps no natural categories for textual features, or at least those features for which we have names are not mutually exclusive categories in a single taxonomy. For example, quotations might

cross paragraph boundaries, or we might have a graphical interest in a hand-written text if we were a graphologist. In summary, in such cases we are faced with 'practical problems which raise philosophical issues' (Biggs & Huitfeldt 1997: 348). These philosophical issues arise because 'there are no facts about a text which are objective in the sense of not being interpretational' (Huitfeldt 1992: 149).

Writing

At the point of production, the empty manuscript page is quite non-directive: one can use it in many different ways. To this extent it is a more liberal environment than sitting at the keyboard. However, the expression of ideas is not necessarily synonymous with the expression of linguistic content. It is only one model of authorial activity that consists of writing/typing linguistic content in a string so that it appears as an equivalent to the printed page. But this one-dimensionality is a technological rather than a conceptual/structural feature. The source may actually contain deletions, over-written amendments, inter-linear additions; it may contain content in a variety of languages, it may contain spelling and grammatical mistakes, it may be simple to understand the content, difficult, impossible for anyone other than the author, or it may be nonsense (impossible for anyone to understand or meaningless). All these are factors that would affect the later interpretation of the notation. In addition the author may use notations such as logic and mathematics, which share with alphabetic writing a set of conventionalized signs but depart from it in the rules for their combination and the spatial distribution of the symbols on the page, for example the content difference between 22 and 2^2; the 'principle of position' (Gelb 1965: 19). Diagrammatic structure can also be exploited in the creation of tables of data which show correspondences according to

spatial relationships of clusters of sequenced symbols. Finally, although this is not an exhaustive list of possibilities, graphics can be introduced which contain non-sequenced graphical marks or tokens. In describing the manuscript source one should also note that all of these possibilities are equally available at any place on the page and there is no requirement to start a new line for a graphic, or to place a table inside a box. There are no necessary devices to signify the beginning of a particular form of notation although there are cues in the structure of the notation itself, which result in confirmation or reinforcement of the hypothesized content.

We should therefore differentiate between markup in which the basic structure is sentential but the principle of position confers meaning on certain elements, e.g. mathematical notation, and fully diagrammatic structures in which there is no sentential organisation at all, such as tables. I will call mathematical (etc.) notations 'distributed-sentential structures' in order to reflect the way in which they occupy a mid-ground in terms of spatial structure. I also propose that, despite being inefficient, such notations may have a linguistic equivalent, i.e. they can be 'read aloud' and converted into sentential structures. Or better: that diagrammatic structures are those which have no linguistic equivalent.

Reading

At the point of consumption the reader is presented with the manuscript page in the absence of the author. Identifying the authorial use of two-dimensional space is much more problematic than the interpretation of orthographic tokens as alphanumeric graphemes. On unruled paper orthographic mistakes and ambiguity in the use of white-space can easily hide the signifying features of the notation and it is only as meaning is

construed from hypothetical interpretations of the notation that we can become confident about a particular reading. If the author is mistaken in his or her use of convention or is being creative then interpretation is further problematized.

The interpretation of the page depends upon the identification of cues or metatextual [pre-textual] elements. In turn this identification is a reflection of the interests of the reader; 'our aim in transcription is not to represent as correctly as possible the originals, but rather to prepare from the original text another text so as to serve as accurately as possible certain interests in the text' (Pichler 1995: 691). The principal deciphering activity is the identification of recognisable letterforms. The reader is cued to seek individual letterforms by the horizontal linear organisation of the marks on the page. Within this linearity the task is to identify letterforms from the variable orthography. Confirmation of linguistic content is achieved by the consistent identification of a character set which itself forms identifiable words delimited by white-spaces and line breaks, etc. At this point the key activity is the suppression of reading individual graphical marks as signifying tokens in preference to the interpretation of these marks as the repeated use of a limited range of graphemes (Coulmas 1990: 51) belonging to a character set. Multi-linguistic texts may add an additional level of complexity but at a word level we are likely to be presented with single character sets at a time. Mixed character sets in a putative word unit (a space delimited string) may signify a non-linguistic notational form such as mathematics or logic. Logic provides an interesting case because examples can be constructed which use the characters from linguistic notation but which do not follow the combinatorial rules of natural language notation.

The cues to a change in language, or language group are to a lesser extent the presence of characteristic letterforms, such as Þ (Unicode 00DE, Latin uppercase Thorn, characteristic of Icelandic and Old English), but to a greater extent by the recognition of linearly organized graphemes delimited by spaces into word-units corresponding to a

natural language vocabulary. Unpacking the activity in this way demonstrates the many processes that precede the identification of linguistic content that forms the default mode of the textual interpretation of source documents. At any of these levels mistakes and ambiguities can interfere with the recognition of authorial content and the interpretation of notation, including the presence or absence of certain content objects such as text or graphics.

All this could be taken as the default mode, in other words, having identified linearly organized graphical marks that comprise orthographic tokens that can be interpreted as graphemes forming words with spatial delimiters, we proceed to interpret the marks on the basis that they continue to be text in the established natural language until we are cued to adopt an alternative interpretational strategy. This corresponds to the assumption in text encoding, that the content consists of Unicode strings unless indicated to the contrary. The departure from this default interpretation is normally cued by a different spatial organisation, or by non-Unicode content. However, the example of typewriter art shows that one does not have to depart from Unicode in order to create graphics. If we describe the normal sentential organisation of the text string (which in English has a left-right in-line progression and a top-bottom block progression), as 'passive' then we can say that 'active' spatial organisation is a cue for an alternative interpretational strategy.

Tables are active spatial layouts of textual content in which the relative position of the elements is itself signifying. They are unlike conventional text in which the relative spatial juxtaposition of elements other than the linear sequencing of the letters and word units, is non-signifying. This accords with Larkin and Simon's definition of diagrammatic structure. Tables are not necessarily indicated by containment in boxes and so it is frequently the lack of expected sentential linearity, or the loss of meaning grammatically or semantically when interpreted as sentential structure, that suggests that the author

may have departed from sentential text-organisational mode. This cues us, on the assumption that the marks continue to be purposive and signifying, to seek other meaningful modes of textual organisation or meaningful non-textual signification. These might be further cued by extrinsic elements such as comments in the preceding text that refer to a table. Finally, an assumption of spatial signification may be imposed as a desperate attempt to satisfy our desire for signification in text that seems sententially disordered or ruptured (the shopping list, jottings, etc).

Interpretation

All this takes place in a context of purpose: an attempt to infer meaning from the source material. What is meaningful will depend on our interests. Perceiving something as meaningful depends on element recognition which in turn requires us to be receptive to a meaningful aspect (Renear 2001: 415). For example, it might be the case that until a graphologist draws our attention to the signification of handwriting, we attribute no particular significance to whether a document is written in the author's own hand or the hand of another. Wittgenstein calls this new awareness 'seeing an aspect' (1953: 213). If we do not even recognize the possibility of signification, e.g. of handwriting, we are said to be 'aspect-blind'. Wittgensteinian aspect-blindness is a factor in the encoding of text. One must 'first recognize the deliberate ambiguity, and then encode it so that the linguistic content and the on-screen presentation preserves these two senses' (Biggs & Huitfeldt 1997: 357). But this depends on seeing the ambiguity in the aspect and this in turn depends upon what interest one has in the text. There are also examples where the presentation is inextricable from the content, e.g. this is underlined, in which 'the medium is the message' (Biggs & Huitfeldt 1997: 356). There are therefore, many contextual

presuppositions to the interpretation of a source and by implication, the differentiation of 'graphical' content from 'linguistic/textual' content.

Conclusion

The reductive method is a useful account of the interpretation of assumed textual content in a source document. Identification proceeds from graphemes to words, etc. Disruption of these inferences causes us to find signification in diagrammatic structures such as tables. An intermediate category of distributed-sentential structures has been proposed, e.g. mathematical and logical notation, which use the principle of position. It is proposed that this category may have 'linguistic equivalent'. It is argued reductively that when content is not sententially organized, nor has linguistic content, then it may be graphical. The advantage of this method is that it can identify graphical content or behaviour rather than relying on the perception of graphical appearance, e.g. drawn lines and pictures. The identification of graphics by the technological resources needed for their reproduction reflects neither the strategy employed in their production nor their consumption (writing and reading). In terms of TEI markup, the <figure> tag as a content descriptor should not be confined to references to external resources, nor should the <text> tag, if considered a content descriptor, be a base tag for a document.

		linguistic content	non-linguistic content
1	sentential structure	<text>	
2	distributed-sentential structure	<notation> (maths, logic, etc)	<notation> (music)
3	diagrammatic structure		<figure> (graphics)

This paper therefore modifies Larkin and Simon's binary description of structure by differentiating 2 from 3, thereby leaving 3 more closely associated with 'graphical

content'. It also implies that the content of a manuscript source is graphical until identified as textual.

The advantage of a reductive approach is to overcome the difficulty of accounting for the enormous diversity of graphical content and therefore of providing a characterising description of graphics in general. By extending the process of document feature description adopted for textual content we can narrow the field in which graphics lie. In particular we can show that sometimes letterforms and other 'textual content' can be used graphically, e.g. tables and typewriter art.

Notes

1 An early version of this paper was read at a seminar at the Humanities Information Technologies Research Centre, University of Bergen, on 21 May 2003. I acknowledge the support of the European Community Access to Research Infrastructures Action and the support of the EU ARI WAB management at the Wittgenstein Archives at the University of Bergen

2 Joint Photographic Experts Group

3 Scalable Vector Graphic

4 :) is a text messaging convention for ☺, signifying pleasure

5 Extensible Markup Language

6 Text Encoding Initiative

References

Biggs, M. (1995) 'Graphical Problems in Wittgenstein's *Nachlaß*' in: Johannessen, K. and Nordenstam, T. (eds.) *Culture and Value: Philosophy and the Cultural Sciences,* 751-761. Die Österreichische Ludwig Wittgenstein Gesellschaft, Kirchberg a/W, Austria

Biggs, M. & C. Huitfeldt (1997) 'Philosophy and Electronic Publishing: theory and metatheory in the development of text encoding' *The Monist* 80(3) July 1997, 348-367.

Coombs, J., A. Renear & S. deRose (1987) 'Markup Systems and the Future of Scholarly Text Processing' [http://www.oasis-open.org/cover/commbs.html]

Coulmas, F. (1990) *The Writing Systems of the World.* Basil Blackwell, Oxford

deRose S. (1990) 'What is a Text, Really?' *Journal of Computing in Higher Education* 1(2), 3-26.

Doblin, J. (1980) 'A Structure for Non-textual Communications' In: Kolers, P., Wrolstad, M. & Bouma, H. (eds.), *Processing of Visible Language*, Vol.2, 89-111. Plenum Press, New York

Gelb, I. (1963) *A Study of Writing.* University of Chicago Press, Chicago

Huitfeldt, C. (1992) 'Multi-Dimensional Texts in a One-Dimensional Medium' in: Henry, P. & a Utaker (eds.) *Wittgenstein and Contemporary Theories of Language.* Working papers from the Wittgenstein Archives at the University of Bergen no.5, 142-161. Bergen, Norway

Larkin, J & H. Simon (1987) 'Why a Diagram is (Sometimes) Worth Ten Thousand Words' *Cognitive Science* 11: 65-99.

Mitchell, W.J.T. (1986) *Iconology.* University of Chicago Press, London

Pichler, A. (1995) 'Transcriptions, Texts and Interpretations' in: Johannessen, K. & T. Nordenstam (eds.) *Culture and Value: philosophy and the cultural sciences*, 690-695. Die Österreichische Ludwig Wittgenstein Gesellschaft, Kirchberg a/W, Austria

Renear, A. (2001) 'The Descriptive/Procedural Distinction is Flawed' *Markup Languages* 2(4), 411-420.

Sperberg-McQueen, C.M., C. Huitfeldt & A. Renear (2000) 'Meaning and Interpretation of Markup' *Markup Languages* 2(3): 215-234.

Sperberg-McQueen, C.M. & L. Burnard (2001) *TEI P4: Guidelines for Electronic Text Encoding and Interchange.* Text Encoding Initiative. [http://www.tei-c.org]

Wittgenstein, L. (1953) *Philosophical Investigations.* Basil Blackwell, Oxford