"Typography is to language what maps are to geography" (Baines & Haslam, 2005, p10)

How has the transition from paper to screen changed the use of typography in cartography, and how can it be improved?

INTRODUCTION TO TYPOGRAPHY

What is typography?

Typography surrounds us in our daily lives. It is something we encounter everywhere we go, and it is something we interact with thousands of times per day. Typography is one of the most important tools that we have for communication, and yet it is something that most of us take for granted. Language, like everything else, is constantly changing (Aitchison, 1981, p16), and so understandably, our relationship with, and our understanding of typography has changed over the years too (Baines & Haslam, 2005, p6). Originally it was the craft of casting type out of metal and arranging it to be printed by hand, a job which was undertaken by few, whereas today, advances in computer technology have seen typography become a digital craft, and one which is no longer restricted to specialists, as type can be created, arranged and printed by anybody with even the most basic computer literacy. There is no one correct definition of typography, and different sources provide numerous different descriptions. However, for the purposes of this investigation, typography is understood as "the mechanical notation and arrangement of language", a definition which was first suggested in Baines' & Haslam's 'Type & Typography' (2005, p7). Whether it's created by hand or by a computer, and whether it's printed on paper, projected on a screen, or viewed on any other medium, essentially, the one element of typography that remains constant, is that it is a method of communication; it is a way of visually representing spoken language. Our ability to communicate through spoken language is what sets us as humans apart from other species, and typography is our way of documenting spoken language, presenting it and preserving it (Baines & Haslam, 2005, p.10; Robinson, 2003, p.36).

The structure of a typeface

Language is essentially a code. It is made up of strings of words put together to convey a message, and equally, words are made up of phonemes which are strung together to convey meaning. These noises, phonemes, are represented by a series of symbols, and it is these symbols which are the bare bones of a typeface. However, words and their meanings are only understood by people who speak the same language, people who are familiar with the code. (Baines & Haslam, 2005, p10; Kane, 2002, p16). For this investigation, the primary focus will be on the Western, or Latin alphabet, and the English language.

The Latin alphabet contains 26 basic characters, which are constructed using strokes. These strokes can be vertical, horizontal, angular or rounded and almost all characters consist of a combination of these (White, 2005, p137), with strokes often meeting at vertexes or creating open or closed counters (Kane, 2002, pp2-4), to create letterforms which are unique and which can be easily distinguished from one another. Despite this, the letterforms of any given typeface do need to have a visual consistency to them, and they do follow the same basic structural rules; they each sit on the same baseline, and they share the same x-height; ascenders and descenders are usually a consistent height too and the line width and the contrast between thick and thin is consistently applied across all the letterforms; the characters have the same serifs or terminals, and they share the same decorative features. It is the design of these differences in shape, proportion and weight which give a typeface it's own style and it's own unique visual identity and personality.

Key principles of typography

When creating a typeface, the designer has to balance form and function; the type needs to be practical and suitable for it's intended use, but equally, it needs to be visually pleasing. Kane (2002, p12) states that typographers have two goals: "easy readability, and an appropriate expression of contemporary aesthetics". In the context of cartography however, the emphasis is usually somewhat different, as first and foremost the aim is for clarity and legibility or information.

Typefaces are classified according to their visual characteristics, the main classifications being serif, sans-serif, slab-serif, blackletter and script faces (Boardley, 2008, online). When setting type, it is important to consider what the type is being used for, as not even the best typefaces are suitable for every occasion (Boardley, 2008, online). Combining different styles of typefaces, and different sizes, can be visually confusing, or, if done appropriately, it can create a visual hierarchy, thus making the information easier to follow and understand.

Another way that a visual hierarchy can be achieved is through the use of different weights; typefaces often have a range of weights from ultra-thin through to ultra-black, and often condensed and extended versions too, as well as italics (Kane, 2002, p8), all of which can be used to create a distinction between different levels of information. The appearance of typography is also affected by the leading and kerning of the type, which control the amount of white space between letters, and this has a large impact on readability.

What makes a typeface legible?

To determine what makes typography legible, we first need to understand how we read text, and how our minds interpret the symbols that make up our written language. In general, we read by recognising the patterns formed by the words; we do not read each and every letter individually, but typically the first and last letters, and the pattern of ups and downs formed by the ascenders, descenders, and neutral characters. Words set entirely in uppercase letters take longer to read than their lowercase equivalents, as they form a rectangular box shape around the word, rather than a recognisable pattern of ascenders, descenders and neutrals. This theory of word shape recognition was first proposed by James Cattell (1886), and his study has since been replicated and expanded upon by numerous others, notably Woodworth (1938), Reicher (1969), Fisher (1975), Haber & Schindler (1981), and Monk & Hulme (1983). It is because our minds cannot interpret uppercase words as quickly as lowercase ones, that uppercase, and also italic letters, are often used to give emphasis to words, as we are forced to read them slower, concentrate on them for longer, and thus pay them more attention (Boardley, 2008, online). Interestingly, it has also been proven that each of us reads differently, depending on the purpose for which we are reading (Unknown, 1927, p124). This is particularly relevant for map reading, as we are not reading lines of continuous, familiar text, but usually the singular, often unfamiliar text of names and places.

Another question that ought to be addressed, is the difference between legibility and readability. Although the two words are often used synonymously, there is a distinction between the two, as discussed by Baines and Haslam:

> "Legibility refers to the typeform, how easy an individual character or alphabet is to recognize when presented in a particular font. Readability encompasses both typeform and arrangement - how easily a text can be read." (2005, p125)

As well as numerous factors that are within the designer's control, the readability of the text is also affected by numerous environmental factors, such

as the lighting, and the type of media that the text is being read from, be it paper, a computer screen, or anything else (Baines & Haslam, 2005, p125). The layout of the text also has a big impact when it comes to readability, and especially is especially important when used in cartography, as large amounts of detail have to be presented in a very small area.

As for legibility, this is affected by the shapes and the proportions of the letters themselves. At small sizes, text can become particularly difficult to recognise, so for maximum legibility the defining features of each character need to be exaggerated to make them easily distinguishable from one another. The amount of white space within and surrounding each letter is an important factor on legibility too (White, 2005, p137) and the counters of letters, particularly three tiered lowercase letters such as e, a, and s, have a tendency to close up at small sizes. The most legible typefaces are those which have a generous x-height to allow for large, open counters, and also a line weight which is neither excessively light, nor excessively bold (Haley, 2001, online). There is also the issue of serifs. Serifs create a noticeable horizontal flow within the text, which many believe helps to guide the eye along the line it is reading (Poole, 2005, online). It has been argued in the past that serifs both improve legibility (De Lange et al., 1993, pp241-248), and that they worsen legibility (Haley, 2001, online). Countless studies have been conducted into the legibility of serif versus sans-serif typography, the most recent being undertaken by Arditi & Cho (2005), but all these studies have yielded results of “no difference” (Poole, 2005, online). Lund (1999) suggests that whilst there may be some slight difference in the legibility of serif and sans-serif typefaces, these differences have so little impact that they are rendered insignificant. What can be agreed on, is that the over-emphasis of serifs will cause a deterioration in legibility (Haley, 2001, online).

How printed type works

The appearance of printed type differs depending on the printing method used. Traditionally, the main methods of printing were letterpress and hot metal typesetting. However, the introduction of offset lithography at the beginning of the 20th century offered an alternative to printers, and although it was very expensive, it was ideal for large print runs. The introduction of phototypesetting machines around the middle of the century made offset lithography more efficient, reliable, and cost effective to set up, and brought about a decline in the use of metal type. Offset lithography is still the most common printing method in use today for commercial printing, and is produced by transferring the text photographically, or digitally, onto a cylindrical metal plate which is chemically treated so that only the image areas will accept ink (Miller, 2009, online). The plate then has oil-based ink and water applied to it, and as oil and

water do not mix, the ink clings to the image areas, and the water clings to the areas that are not to be printed (Bear, 2009, online). Finally, the ink is transferred to an offset cylinder before being pressed against the paper (Miller, 2009, online; Bear, 2009, online). Increasingly, people are using small-scale personal printers for home and office use, and nowadays, inkjet and laser printers are commonplace. However, these bring with them numerous complications, as type is displayed differently on screen to how it is on paper. Most digital type is optimised for screen use, and when output to paper the shapes and colours often appear different to how they look on the screen (Brown, 1993, p134; Gardner, 2006, online). These complications will be further explored later in this investigation, focussing on cartography and the issue of printing digital maps onto paper.

How digital type works

Digital type works in a completely different way to printed type. When viewing printed type, the eye is looking at a series of marks which have been imprinted onto paper, whereas when viewing digital type, the eye is looking at a series of pixels which are being displayed on a screen. Understandably, the quality and the sharpness of digital type is dependent on the quality of the screen that it is being viewed on, with the most important factor being the resolution of the screen (Tam, 2006, p6). The majority of contemporary computer screens use a Liquid Crystal Display (LCD) rather than the older Cathode Ray Tube (CRT) system.

In order to understand how digital type is rendered on a screen, it is first necessary to understand the basics of how LCD screens work. An LCD display is made up of a grid of pixels, and each pixel is capable of emitting light, so that when combined, the pixels can form images on the screen. Each individual pixel is made up of a row of three sub-pixels, and each of these sub-pixels has a colour filter applied to it, to create the colours red, green and blue (Tyson, 2000, online). The amount of voltage applied to each sub-pixel controls the intensity of it's colour, and each sub-pixel has a range of 256 different shades (Tyson, 2000, online), which can combine to create 16.8 million different colours. It is worth noting that whereas screen displays use combinations of red, green and blue to create all the other colours of the spectrum, printed material combines colours in a different way, mixing cyan, magenta and yellow to create other colours; this is known as subtractive colour mixing, whereas the RGB method is additive colour mixing (Van Holten, 2005, online).

LCD screens come in a variety of sizes and applications, but since the year 2000 almost all LCD computer screens have had a resolution of 96 pixels per inch (PPI) (Knight, 2003, online). This is remarkably small in comparison to the

resolution of printed matter, which is typically around 2,500 dots per inch (Tam, 2006, p6). Printed matter is measured in dots per inch, rather than pixels per inch, although the units are essentially the same. A typical inkjet printer is capable of producing 300-600 DPI, and commercial printers are capable of far higher resolution. It is suggested that any detail beyond 300 DPI is relatively undetectable to the unaided human eye, but below 300 DPI, differences in quality do become particularly noticeable (Wikipedia, 2009).

A digital typeface is stored on a computer as a set of 'Bezier curves' or 'quadratic curves', and this data tells the computer about the outline shape of each letter (Baines & Haslam, 2005). These curves are stored as a set of equations, and so unlike images which are saved as a set of pixels, the letters can be scaled to any size without any loss of quality. However, even though digital typefaces are not stored as pixels, they have to be displayed using pixels due to the nature of the screen display; the Bezier curves tell the computer what shape the letter is, and the computer uses this information to determine which pixels to turn off and which to turn on in order to render the letter (Tam, 2006, pp2-7). When type is displayed as pixels this way it is known as a bitmap. Due to the relatively low resolution of screen displays, the curved parts of the letterform often appear to have jagged edges, because there are not enough pixels in the bitmap to create a smooth curve, and so the crispness and the clarity of the type is compromised. Rubinstein referred to these aliased bitmap letterforms as "Jaggies" (1988, p48). The way typographers get around this problem is by using a technique known as anti-aliasing. Anti-aliasing works by changing the pixels immediately surrounding each character to varying shades of grey (Tam, 2006, p7), and this gives the appearance that the outlines of the type gradually fade from black to white. Instead of the pixels simply being on or off, like they are with aliased type, anti-aliasing provides the option for pixels to be somewhere in-between the two states, with varying intensities of brightness, and this creates the illusion of smooth curves for the letterforms.

The downside of anti-aliasing is that the varying shades of grey make the type appear less black than it did before, and although not a problem for display type, at small sizes this can make the text look very faint, and in some cases almost completely illegible. For particularly small text, say 8 point or below, it is often best to turn anti-aliasing off, and instead typographers use a feature known as hinting. Hinting is the process of individually adjusting the pixels of each character to provide the best possible bitmap image for the type at small sizes (Tam, 2006, p8). It is necessary for hinting to be done manually, because the automatic bitmap image determined by the computer may not always be the most legible.

INTRODUCTION TO CARTOGRAPHY

What is cartography?

Cartography, just like typography, is constantly evolving, and can mean different things to different people. Maps are made for a whole range of different purposes, from navigating an unknown environment, to visualising population statistics, or the temperature of ocean currents. However, just like typography, the primary function of cartography is to visually communicate information. I cannot find a more appropriate description than that given by the British Cartographic Society:

> "The function common to all maps is to show selected geographical features of the world in the clearest possible way." (Darkes & Spence, 2008, p13)

Key principles of cartography

When designing any map, the designer is faced with the challenge of representing the three dimensional world on a two dimensional plane (Keates, 1996, p84; Pitkänen, 1980, p42; Meng, 2005, p5). Typically, this is achieved by representing the three dimensional world through a series of lines and symbols. The British Cartographic Society identify three main constituents that make up any map (Darkes & Spence, 2008, pp26-27):

- *points*, meaning symbols, icons, and semiotics in general;
- *lines*, such as roads and borders;
- and *areas*, which form the background of the map, and are usually coloured or patterned.

However, as typography is central to this investigation, it is proposed that *lines*, *symbols* and *typography* might be a more appropriate explanation of what comprises a map. Not all maps include typography, and there are many thematic maps which don't (Dent, 1999, pp113-131), but for all the maps relevant to this investigation, it is from these three basic elements that each is formed. When designing a map, the cartographer manipulates these three building blocks of lines, symbols and typography, varying their "size, shape, colour, lightness, orientation, and pattern or texture" (Darkes & Spence, 2008, p26), to give the map meaning, as well as it's own style and appearance. However, these are purely cosmetic factors contributing to the appearance of the finished map; the appearance of the map will also, inevitably, be dictated by it's purpose (Lee, 1995, p36). According to Lee, the most important consideration for any cartographer, is "what is the map to be used for?" (1995, p36). Lee's work is expanded upon by Darkes & Spence, who say that every

map must be designed for a specific purpose, with a clear and definite end use and target audience in mind (2008, p36), as well as considering the money and equipment available to make the map (2008, p18).

Cartography is concerned with map making of all varieties, and maps are made for a huge range of different purposes (Darkes & Spence, 2008, p13). This investigation will only be concerned with topographic general purpose maps, often referred to as road atlases or reference maps, as these are the most common maps in use today, and have made the most significant impact on the world of digital cartography, hence they are the most relevant. Robinson & Petchenik define reference maps as those which are primarily concerned with navigation and location, and which provide information about the geographical features of a certain area (1976, pp116-117).

What makes a map legible?

For any map to be an effective and easily understandable piece of communication, the cartographer must decide which geographical elements are worth including and which are worth omitting. “A map can't show every feature of the landscape and should never try to” (Darkes & Spence, 2008, p22). Soffer et al.’s research into the legibility of cartography for use in atlases showed that “a high density of information is counter productive” (2005, p5) and “reduces the legibility of maps to minimal levels” (2005, p5). Interestingly, the search and identification tests conducted by Soffer et al. also suggest that when concerning map legibility, the shape of map elements is more important than their size or colour (2005, p5). However, they note that this result seems counter-intuitive, and admit that it could be an anomaly due to the small size of the sample taken (2005, p5).

The British Cartographic Society note that for a map to communicate effectively, first and foremost the data used to create the map must be of a high quality. Geographical source data must be “complete, consistent, and accurate” if a reliable and honest map is to be produced (Darkes & Spence, 2008, p31). A strong visual hierarchy is also very important, as it helps the reader to quickly interpret the map and pick out which elements are relevant to them and which are not. Typically, the larger and darker the feature is on the map, the more important it is (Darkes & Spence, 2008, p34). Elements such as colour and visual contrast, if used effectively, can also help users to easily distinguish between different geographical features (Darkes & Spence, 2008, p34), with natural and true colours, understandably being proven the most recognisable (Soffer et al., 2005, p6). Techniques such as simplification, exaggeration, and displacement, are also used by cartographers to make maps clearer and easier to understand (Darkes & Spence, 2008, pp22-23).

How people interact with maps

XYZ