# Following print: What does this really mean in the electronic environment.

Kathy Riessen.

Coordinator Accessible Format Production, South Australian School for Vision Impaired.

Executive member, Australian Braille Authority

## Background

As a sighted transcriber with the Department for Education in South Australia, I transcribe into braille literary, mathematics and other technical material, music, and languages such as French, German, Italian, Indonesian and Japanese. As technology has advanced and with the adoption of Unified English Braille, in the 30 plus years of my career, I have seen many changes.

One of the biggest changes that has occurred in the past 30 years is being able to source electronic documents to speed up the production of braille.

When I started my career as a transcriber, I would be given a paper document or book to painstakingly reproduce into braille using a Perkins Brailler. I was able to reproduce the text according to what my eyes and brain interpreted the text was saying. Then came the advent of the personal computer with word processing software. As these became more available braille translation software was developed which could take an electronic document and translate it into braille from which a hard copy could be made using an embosser. I could take my printed document and manually create an electronic document either typing it directly or scanning and editing. This document would then be put through translation software, edited and embossed. Again, as I created the word processing document, I was able to ensure that what I read with my eyes on the page was faithfully reproduced into braille.

The next technological development was OCR software whereby a printed page could be scanned and converted to electronic text, eliminating the need for manual typing. Around the same time came the braille display whereby not all braille needed to be physically embossed but could be read directly from a computer.

Now, more often than not, our source document comes to us already in an electronic format whether it is from a classroom teacher or a publisher. I now rarely need to manually type or scan text to produce braille.

Over the years, the accuracy of braille translation software has undergone continual improvement to the extent that the accuracy of translation is extremely high including mathematical and technical material.

But … the question needs to be asked, “is an electronic document the same as a printed document?” How can we ensure that an electronic document when translated to braille to be either physically embossed or read via a braille display, will accurately reflect what the original document says?

My personal experience is that whilst this is mostly correct, this assumption cannot be made.

## Reading print

When I read a page of print or a page on a screen using my eyes, they automatically interpret the page taking into account the context of the surrounding text.

But is what my eyes read and interpret the same as what the electronic document actually says. What do I mean by this? To answer this we need to look at four major concepts.

## 1. Typographical symbols

The first area to look at is typographical symbols. There are many symbols used in both text and technical material which although look similar, are not the same. Two symbols may look similar but have a very different meaning, and have a different braille equivalent.

The standard QWERTY keyboard has a limited range of symbols, but word processing allows us to produce nearly every typographical symbol possible using its UNICODE name and number. UNICODE is an international consortium which has developed a way by which every possible symbol can be produced, and most translation software will correctly translate many UNICODE characters.

However, many sighted producers of texts, are lazy and work on the premise that “if it looks right it must be right.” As long as the eye can interpret a symbol it is OK to use it even if the symbol used technically has another meaning. I have found this approach to be true with not only day-to-day material produced by a classroom teacher but also in published texts.

The following are some common examples I have come across where incorrect symbols have been used in electronic documents. They all read perfectly well visually, but as the specific character has not been used, they will be incorrect when translated into braille. Some of these have greater impact on the readability of braille than others. A good braille reader can often cope with some of these incorrect interpretations using their background knowledge of context to enable the reader to interpret the writer's intention. This does not however excuse the incorrect use of symbols. I work in the education sector, and it is important that students, particularly at the lower levels are given braille which is correct.

Example 1:

26°C, 26oC

#bf^j,c1 #bf9o,c

When read visually both of these examples read as 26 degrees Celsius. The degree sign is a small circle in the superscript position. The first example has used the correct symbol, UNICODE 00b0, but the second uses a lower case letter o in the superscript position.

These two symbols look very similar, can be visually interpreted the same within the context of immediately following a number, but when translated into braille produce two very different results.

Example 2:

4×5, 4x5

#d"8#e1 #dx#e

When read visually, these two examples will read as 4 multiplied by 5. The multiplication sign is a small cross very similar to a lower case x. Because they look similar the lower case x is often used instead of the correct symbol, UNICODE 00d7, particularly by teachers producing worksheets for students as it is quicker to type.

On the printed page there is no confusion that this is a multiplication, but when translated to braille, the difference is clear, which is unacceptable particularly for young students.

Example 3:

«Hello», ≪Hello≫

\_8,hello\_01 .@<,hello.@>

The context of these examples is a word within angled quotation marks or guillemets. The two examples are visually similar and so to the eyes read will read and interpret these as quotation marks. However the first uses the correct angle quotation marks or guillemets, UNICODE 00ab and 00bb, where the second uses the “much less than” and “much greater than” symbols, UNICODE 226a and 226b.

The opposite effect occurs where angled quotes are used in the context of much greater than or much less than.

Example 4:

"Hello", ''Hello''

8,hello01 '',hello''

When read visually, both of these read as a word in double quotes. However in the second example two straight single quotes have been used to create what looks like a double quote, which then translate as two apostrophes.

I have found this to sometimes occur in newspaper articles or electronic texts from sources such as Project Gutenberg.

Depending on the quality of the original print, scanned text which is then put through OCR software will often have errors such as:

* Misinterpretation between 0 (zero) and O (capital letter o)
* Misinterpretation between 1, lower case letter l, and capital I.
* m being interpreted as rn, and many other examples.

How quotes and apostrophes are represented in electronic texts can also have a bearing on whether or not braille translation reflects the print meaning.

The single quote and the apostrophe are usually represented in print using the same symbol, however they have different braille representation. As a sighted print reader, I determine via context the difference between a quote and an apostrophe. A good braille reader can also do this if the wrong braille symbol is used. However, within the education sector it is important for students to learn and read the correct braille symbols.

In Australia the most common typographical approach is to use single quotes as the dominant quote. These are most commonly represented using angled quotation marks or in Word “smart quotes”. The apostrophe whether within a word or at either end of a word is represented using a closing single quote. Translation software will usually translate this character in the middle of a word as an apostrophe. When positioned at either end of a word it is usually translated as a single quote, unless the software knows the that word is a common word which begins or end with an apostrophe.

I am not going to go into detail in this document how to ensure correct translation of quotes and apostrophes, except to say that my experience has been that to ensure correct translation when using the Duxbury Braille Translator, you can use the fact that a straight apostrophe, UNICODE 0027 will always translate as an apostrophe, and an angled single quote at the beginning and end of a word will translate as a single quote.

In summary, to ensure that translated braille correctly reflects the original print, the correct typographical symbols need to be used.

## 2. Heading structure

As a sighted person, when I read a print document, be it a textbook, newsletter, worksheet, etc., a quick visual scan of a page can give a lot of information. How a page is laid out, particularly in terms of headings gives the reader clear information about how the document is constructed.

An example is a restaurant menu which may be a number of pages long. I don't need to read every word from the beginning to find where the main courses are. I can quickly skim past the headings “Soup” and “Entrée” to find the heading “Mains”. Similarly, I don't need to read the whole wine list to find a good red wine. I can simply find the “Red Wine” heading. To reflect print, well formatted braille documents similarly include clearly defined headings to aid quick navigation. Whilst each country has differing protocols of how headings are to be formatted in braille, the importance of headings is universally recognised.

There are two ways in which electronic documents are constructed.

* Using styles
* Manually formatting each paragraph

When an editor constructs a document using styles, translation software is able to then map these across to the translated document and format the headings according to the braille heading protocols required.

Unfortunately, most document editors are either lazy or unaware of how styles may be used and manually format each paragraph to create the document structure. They work on the premise of “if it looks like a heading, then it must be a heading.”.

Headings stand out in print as they are formatting differently to the surrounding text using devices such as a larger and bold font, different colour, different font, etc. Headings stand out in braille by means of the use of blank lines and specific positioning on a braille line. Bold or italic typeforms are not usually required in braille headings even if these occur in print.

A document which has been manually formatted, when translated will simply be a series of paragraphs. There is no structure to the document and any bold or italics which form part of the print heading will be shown which makes the document less navigable.

Whilst an electronic document which has been produced using styles, may look and read the same to a sighted reader as the same document produced with manual formatting, only the former can be deemed to accurately follow print.

## 3. Reading order

The layout of many modern textbooks can be quite complex with multiple columns, interspersed with small sections of side information, tables, diagrams, charts etc. A sighted reader visually scans a page and determines at a glance a logical order in which to read the information, keeping relevant information “together” as it is read, or skipping over sections which may break the logical flow of reading.

One of my roles of a transcriber is to ensure that braille is produced in a logical linear reading order to maintain the best reading flow of information.

An electronic document also has a built-in reading order, which is used to determine the order that text is translated using translation software. Depending on the complexity of the document, this may or may not be the same as the logical reading flow.

I have found that most documents produced using a standard word processor such as Word already have a linear reading order. However, when a document contains columns, unless they were created correctly, the reading order may be disjointed. I have come across numerous instances where the editor of a document rather than defining columns in the layout feature of the word processor, has used tabs to create columns which are to be read down. In this instance the reading order is “first line of the first column”, “first line of the second column”, “second line of the first column”, “second line of second column”, etc, rather than reading all of the first column then all of the second column. A similar problem may also occur where OCR gets confused with columns in a scanned document.

Textbook PDFs, which are a common source of document for us, do not always follow a logical reading order. A good PDF will have the reading order defined, but these are few and far between. When constructing these documents publishers are more concerned with how the final outcome looks, rather than how it reads electronically. With these PDFs when reading the text, either using a braille display or copying and pasting into a word processor for translation, the logical flow of information is lost unless the reading order was correctly defined at construction.

How an electronic document is constructed with the view of reading flow or reading order determines whether or not a braille document created from it accurately reflects the original print.

## 4. Graphical items

Reading print does not just involve reading the printed word but also interpreting any charts, diagrams, pictures etc. As a sighted person reads a page, they are also absorbing information from graphical information on the page.

In order to fully reflect what is on a printed page, the information from non-text items also needs to be included in the braille if it is to fully reflect the print page.

This paper is not going to go into detail on how to ensure these items can be included when translated into braille, however at a minimum the following two points help with ensuring that the information from graphical elements are included:

* Graphical items should be correctly positioned within the proper “reading order”
* Graphical items should have an alt-text caption which gives the relevant information about the graphic to the reader.

The reality is that most document editors have no concept that the information in graphical elements would be of interest to a blind person who may want to access the document.

Electronic documents which contain graphical elements do not fully reflect the original print unless appropriate reference to those elements is made.

## Summary

The digital age has arrived. This is an exciting time and there is now access to an unprecedented level of written information.

With the aid of translation software, braille displays and embossers, a formal transcriber is no longer always required for a blind person to independently access information.

Some people however have the misguided concept that this is sufficient. And whilst it is in many cases, it does not diminish the role of a good editor or transcriber to ensure that all the information from a print or screen page is accurate and can be accessed by the braille reader.

The following checklist should be applied to any electronic document before it can be accurately translated:

* Have the correct symbols been used? Are quotes and apostrophes correctly represented?
* Have styles been correctly used to create a heading structure?
* What is the reading order? This can be determined in a number of ways. Move the cursor through the document and watch the order in which it moves or reading with a screen reader.
* Are there any graphical elements? Are they correctly positioned in the document and do they have alt-text captions?

The question was asked at the beginning “is an electronic document the same as a printed document?”

My experience says that this assumption cannot be made. There is still a lot of education and work to be done to get to this point. Most electronic documents are produced by print readers who have no concept of braille or accessibility and so we get the errors discussed above and more.

In my own little sphere of the world, we try and encourage teachers and aids preparing worksheets to think about the above points with varying success. In an ideal world, editors, publishers, teachers would be taught how to construct documents properly from the very first time they use a word-processer at school, but this is a long way off.

In the meantime, transcribers around the world are doing their very best to ensure that the braille they produce, accurately reflects the original print, even if the original electronic document they are working from does not.