

## REPRODUCING *THE WOMAN IN WHITE* IN ORIGINAL PARTS

There are several ways to convert printed text into a computer file. One, of course, is for a human to copy-type it at the keyboard. Fortunately that is no longer necessary as the computer can do most of the work.

### **I. E-texting**

The original page is scanned in as an image. Software then converts that image into characters in a text file. The accuracy of this software is now amazingly good. But there are inevitably mistakes, especially when working from copies of 19<sup>th</sup> century originals, and these have to be corrected. This first part of the job is being done by James Rusk in Texas. He writes:

My method was refined by several years of contributing texts to Project Gutenberg. Using a standard scanner, I scan the image and then use an OCR (Optical Character Recognition) program to convert the image to text. My favorite OCR program is ABBYY FineReader Pro, which does an excellent job and can scan text from many different languages.

Once the scanning is done, I use Word to format and correct obvious errors. Spell checking helps spot cases where, for example, an “e” was read as a “c”. For the final edit, I print the pages and read them slowly with red pencil in hand. Some people can do the editing on screen, but I find I catch more errors by reading a printed page.

As a result of that care, the files James sends are more than 99% accurate. But there are still several stages to make sure that the text is as close as humanly possible to the original.

### **II. Formatting**

First, I use search and replace to convert certain symbols. For example the text file will contain two hyphens -- instead of a long dash —. Unsexed quotation marks and apostrophes like these "" '' are converted to their sexed equivalents “ ” and ‘ ’. Any double spaces and double paragraphs are converted to singles. Paragraphing has all automatic features turned off except that the first line of each paragraph is indented by 1 cm. The font is converted to Century Schoolbook and the format is adjusted to fully justified – straight left and right margins. Italic text is marked for manual conversion to italics. Most of this is now done by a Word macro which records and replays formatting keystrokes.

### **III. Read aloud proof with punctuation**

I then proof read the text myself using another piece of software called Natural Reader. It reads text aloud with very realistic voices – male or female – and with an English or an American accent. I tend to use the English Charles who is almost as good as having a human read it to you! The software has some eccentricities. For example ‘waistcoat’, which I would pronounce as it is spelled ‘waist’ ‘coat’, the software pronounces ‘wesskit’. And occasionally it gets a word wrong or, more often, pronounces it for the wrong context—‘close’ and ‘minute’ being good examples. Some words such as ‘learnt’ and ‘learned’ are hard to distinguish and ‘connection’ must always be changed to its Victorian sound-alike ‘connexion’. Recently I

discovered that the pronunciation of ‘crochet’ and ‘crotchet’ are almost indistinguishable.

The intonation of the voice gives very strong clues to the punctuation – the presence of a comma causes a pause, for example. But I found it was not accurate enough to pick up the eccentric punctuation in *All The Year Round* and errors – especially the presence or absence of commas or semi-colons – were being missed.

At first I went through the file using search and replace to convert all the punctuation marks to words. That was very time consuming so I devised a macro in Word to do the work for me. Now if I press Ctrl Alt Z all punctuation is converted to words. So this paragraph...

“You quite understand, Sir Percival?” the lawyer went on.  
“Lady Glyde is to sign her name in the presence of a witness—or of two witnesses, if you wish to be particularly careful—and is then to put her finger on the seal, and say, ‘I deliver this as my act and deed.’ If that is done in a week’s time, the arrangement will be perfectly successful, and the anxiety will be all over. If not——”

...becomes...

[DOUBLE QUOTE]”You quite understand [COMMA], Sir Percival [QUESTION MARK]? [DOUBLE QUOTE]” the lawyer went on [STOP]. [DOUBLE QUOTE]”Lady Glyde is to sign her name in the presence of a witness [LONG DASH]—or of two witnesses [COMMA], if you wish to be particularly careful [LONG DASH]—and is then to put her finger on the seal [COMMA], and say [COMMA], [APOSTROPHE]’I deliver this as my act and deed [STOP]. [APOSTROPHE]’ If that is done in a week [APOSTROPHE]’s time [COMMA], the arrangement will be perfectly successful [COMMA], and the anxiety will be all over [STOP]. If not [LONG DASH]— [LONG DASH]— [DOUBLE QUOTE]” [PARAGRAPH]

I do this expansion in a copy of the file and then get the software to read the expanded text aloud. I check it against the original in the contemporary volumes of *All The Year Round* using a ruler to make sure my eye does not wander from the line. When I hear a discrepancy I stop the reading and check both texts. If it is an error I correct it and highlight the correction.

Once the part has been read through, I correct the original file from the highlighted errors. Italic font is inserted where indicated. I do a final check to make sure that all the changes have been made.

That file is then copied into a Word template with two columns and the image of the heading at the top and saved. I do not look at it again until the week it is due to go out.

As I proof read the file I note any apparent printing errors. Some are more obvious than others. Any suspected errors are compared with the first edition of the book published in three volumes in 1860 and with the first one volume text published in 1861. I also consult the Oxford English Dictionary to check on spelling or usage. I then add the notes to the errors file.

#### **IV. Final proof and layout**

In the week the Part is due to go out – usually on Wednesday or Thursday – I proof read it one more time using the same technique: copying the text into another file, expanding punctuation, and using the reader. Usually that reveals one or two remaining errors. Perhaps I should read it another time until no error is found. But I do not.

Once corrected, the two column Word file is checked carefully for layout, line and page breaks, capitalisation and centring. The correct part number and date are inserted at the start and end of the text and in the header and footer. The correct page number is inserted on the first page. When I am completely happy with the look of it a pdf is created using Adobe Acrobat. The web page for the part is created using the same text.

Then I scan in the original *All The Year Round* pages and the illustrations from *Harper's Weekly*. Those are converted to pdfs

I write the introduction for the weekly email, create the remaining web pages and copy the pdfs over to the web pages on my computer.

#### **V. Publish**

Early on publication day I upload the new and revised files to the website, check they all work as they should and the links are correct. Then I send out the email to the nearly 550 readers worldwide. Half an hour later I send it again through a different route to half a dozen that are inevitably returned by spam filters.

Paul Lewis  
8 March 2010