

10 — HOW TO KEYBOARD

PDFs

Portable Document Files (PDFs) are files with .pdf file-extension. They may contain text or images or both. In WEBBS we send out work to be keyboarded as PDFs containing only images — images of pages of an original Bible or other document. Our work is normally sent out in the form of PDFs as e-mail attachments.

How Much to Keyboard

It is important to type only the chapters allocated to you by your team leader. The length of each PDF is usually eight pages of the original text and this does not usually amount to a whole number of Bible chapters. For example, a PDF which has a filename bemMAT01-03a.pdf would contain about 2½ chapters, and the following one with filename bemMAT03b-05a.pdf would contain about 2 chapters.

So, if the work allocated to you is Matthew 1–3 both those PDFs will be sent to you but you should not type beyond the end of chapter 3. The chapters after that point will have been allocated to someone else.

Be careful to type only those cross-references and/or footnotes (if there are any) associated with the chapters allocated to you.

Handling PDFs on your Screen

Rotating: When you open a PDF you may find that the text is sideways on the screen. Click on the “View” menu of your PDF reader program and then, depending on what version of reader you have, you will see how to rotate the file — clockwise or anticlockwise — to put it the right way up on your screen.

Enlarging a detail: One of the benefits of looking at these files on screen instead of on a paper copy is that you can enlarge a small area to see small details. This can be done by the zoom method or, if you have a mouse, by Ctrl, + mouse wheel.

If you use Adobe Acrobat Reader you can also open the pdf and the InSheet within EasyKeyEdit4. Select File: Open pdf file to keyfile and/or View Instruction Sheet. There is more information in Help/how to.

[Note: this feature will not work with other pdf readers, but you can download Adobe Acrobat Reader for free from [Acrobat Reader \(easy-key.info\)](http://Acrobat Reader (easy-key.info)) or from <https://helpx.adobe.com/ca/acrobat/kb/install-reader-dc-windows.html>]

Checking after Keyboarding

Although careful checking of each line as it is finished may remove all actual typing errors, other mistakes may remain. It is essential to check after the whole file has been keyboarded for such things as errors in SFMs. Checking the file for these and other errors may best be done using Easy-KeyEdit4 Preview option.

Checking by using Easy-KeyEdit4 Preview

When you have completed and proof read your file, it will also be helpful to check the file using Preview, which will show you how the file might appear when it is formatted ready for publishing. While it is not a substitute for thorough proofreading, this view may also point out some mistakes you might have made with SFMs, and provide an alternative view for proofreading.

If a correction is needed, this has to be done in Text Entry as Preview is read-only.

Returning Files

After checking the files for mistakes you should send them to the Team Leader within one month. If you cannot do it within that time please tell your Team Leader. (The arrangements are a little different in countries other than the UK.)

A Note About OCR

Optical Character Recognition (OCR) means the processing with specialised software of printed characters such as those in a PDF, or else from a printed sheet by using an optical scanner. The computer recognises the letters and other characters and produces a text file. It has a similar effect to keyboarding the text from the file. (A PDF, Portable Document File, is a file of a type which has a .pdf file-extension. The ones we use are pictures of pages with text on them.)

We do not recommend OCR for the reasons set out below. If it is used then the precautionary advice must be followed. If, as a keyboarder, you do use it, then please tell us by adding `OCR` to the `\id2` line of each file for which you used it, as:

```
\id2 [language, 3 digit language code within round brackets], [your name], OCR
```

OCR is excellent for some purposes, like large and clear printed text in English with a simple layout on the page, but there are some real pitfalls which, from our experience of having to put them right later, should be avoided by careful sight-checking of the scanned text.

Some people who have an OCR program on their computer may prefer to use OCR rather than keyboarding our files. After doing that it is of course necessary to work through the file putting in the SFMs, and it is also necessary to check very thoroughly for errors and to correct all that are found.

OCR usually produces numerous errors. Common ones when dealing with English script include confusing I (upper-case i) with l (lower-case L) and 1 (numeral one); confusing O (letter) with 0 (numeral); reading letter m as rn; and many more. In the non-English languages that we keyboard there is the additional problem of special characters and diacritics which OCR cannot recognise.

The errors produced by OCR are often difficult to recognise by eye. For example, `walk` (with numeral one as its third character) is hard to distinguish from `walk` (correctly spelt). Another example is that letter o with acute accent, `ó`, may be read by OCR as numeral 6, which in the fonts that we need to use may be difficult to detect.

If two keyboarders doing the same batch of text both use OCR, then the same errors are likely to be common in both files. If they remain uncorrected, when it comes to merging and checking the files such 'matching' errors will not be picked up by the merging software. This means they may remain wrong in the final files.

OCR may take less time to 'capture' text than keyboarding does but this advantage is offset by the longer time it takes first to insert the SFMs and then to do the extra thorough checking and correcting that is necessary. This may still save time for the keyboarder but it is our experience that some files produced by OCR have many more errors remaining when they are sent in than keyboarded files do, and many are types of error that are more difficult to detect. That causes delays further down the line and there is not likely to be any overall saving of time.

For these reasons we do not recommend the use of OCR for our work.