

# PDF Files and Translation

*By Jamie M. Lucero*

**As Adobe Portable Document Format (PDF) files** become increasingly widespread for various forms of documentation and easier to generate, translators and agencies are regularly asked—and sometimes expected—to handle PDF source files. Before examining why such requests can be problematic, it would probably be useful to explain why PDFs are so popular.

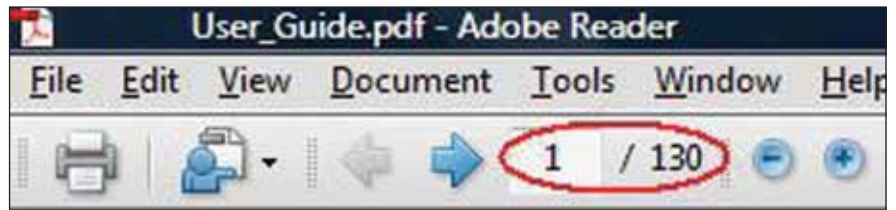
As the name of the format suggests, PDFs are intended for information sharing, much the same way as paper documents. Like hardcopy, a PDF enables a text to be viewed with all the

original design elements in place, thus avoiding the scenario where a beautiful layout is displayed incorrectly on another computer. Because the formatting can be displayed correctly on the screen, it is not necessary to print the file. Since PDFs can be distributed via e-mail, they are also convenient. As an added bonus, they offer protected content that allows text searching, which is definitely beneficial to both authors and readers. Files can also be viewed with free software such as Adobe Reader.

What, then, is the problem? PDF documents—with the exception of forms—are meant for viewing, not

modification. Similar to handling hardcopy source texts, a PDF source file cannot be used directly as the basis for the translation. A new file must be created in order to perform the translation, which creates extra work and challenges for any translator or agency. Add to this the sheer number of software applications that can quickly generate PDF files, and you have the makings of many a translation nightmare. However, with the proper strategies for handling PDF files in place, this much-maligned source file format can be palatable, if not profitable.

Figure 1: Adobe Reader Toolbar



## Translating PDFs

There are essentially two strategies one can use to translate PDF source files: 1) create a new blank file in the program where the translation will be performed; or 2) convert the PDF source file to another format where the source is overwritten with the translation. For the purposes of this article, let us assume that the new format is Microsoft Word.

### Strategy 1: Create a Blank File

This approach is most effective when dealing with small PDFs of one to two pages of text containing minimal formatting. It is an approach that can be used by translators and agencies that do not use translation memory (TM) or related tools. Just follow these steps:

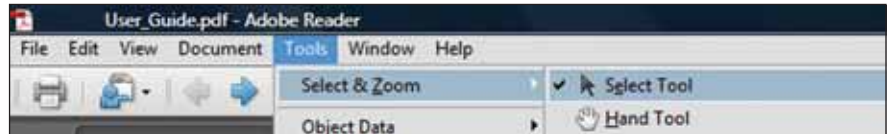
1. Create a new blank document in MS Word.
2. Type the translation into the new document while viewing the PDF (on screen or a printout).
3. Re-create the formatting of the original PDF in the new document according to the specifications agreed upon with the client. (This can be done during or after Step 2.)

The advantage of this approach is that translation can start immediately with little or no file preparation. The flipside is that all of the required original formatting has to be completely re-created in the new document, and the chances of omitting text become increasingly high with larger files.

### Strategy 2: Convert the PDF Source File

This approach is most effective when working with larger PDFs of

Figure 2: Adobe Select Tool



## PDFs offer benefits to both authors and readers, such as protected content that allows text searching.

more than two pages of text that are heavily formatted. It is a good option for translators and agencies that do use TM or related tools. Just follow these steps:

1. Convert the PDF file to MS Word.
2. Fine-tune the formatting in the conversion according to the specifications agreed upon with the client.
3. Translate using the preferred TM tool.

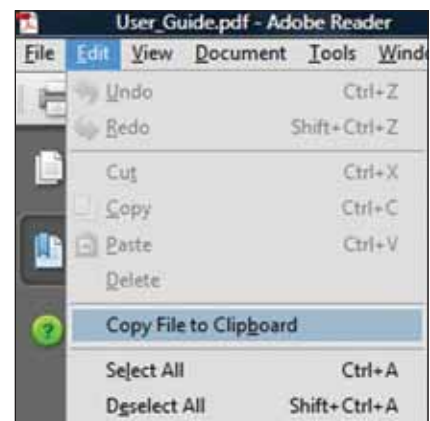
The advantage of this approach is that much if not all of the necessary original formatting can be transferred during the conversion to MS Word. In addition, TM tools can be used, and the chances of omitting text are small, regardless of the size of the PDF. On the other hand, purchasing third-party conversion tools could be required in some cases. The time needed to prepare the document is also much longer. This longer preparation time, however, is frequently offset by the major production gains obtained from the ability to use a TM tool on large documents.

### PDF-to-Word: Conversion Assessment

Before attempting any kind of conversion, it is advisable to assess whether the PDF will be fairly easy to convert. To do this, open the PDF in Adobe Reader and look for the following:

1. The number of pages, images, tables, and charts. The total number of pages is listed on the toolbar, but the other items must be counted manually. (See Figure 1.)

Figure 3: Adobe Reader Clipboard



2. Complex images, tables, and charts. Complexity includes things like the amount of information in these images, non-standard placement of images and charts (e.g., in the margins), text within an image, and large graphic sizes.
3. Heavy formatting, including the liberal use of colors, fonts, and columns.
4. Non-selectable text. To determine this:
  - a. Go to the Tools menu, select Select & Zoom and click on Select Tool. (See Figure 2 on page 17.)
  - b. Try selecting any text in the file to ensure that the PDF is not a scanned document. If you can select text, then it should be available for copying and pasting.
5. Document restrictions. To determine this:
  - a. Check to see if “(SECURED)” follows the file name in the Adobe Reader title bar at the top-left of the Adobe Reader window. If so:
  - b. Go to the Edit menu and verify that the Copy File to Clipboard option is available. (See Figure 3 on page 17.)
  - c. If not, the PDF is protected against copying, and requires one of the following options:
    - Ask the client to remove the security.

- Purchase and use a third-party security removal tool.
- Print the PDF (if permitted) and scan the printed document with optical character recognition software to an MS Word file.

Any of these five assessment items alone or in combination increases the difficulty and time required for conversion. Performing such an assessment will no doubt save some initial headaches, and may even help determine if a given project can be done or is worth doing.

### PDF-to-Word: Manual Conversion

A manual conversion approach assumes that no software tool for such a conversion is available. To perform a manual conversion, do the following:

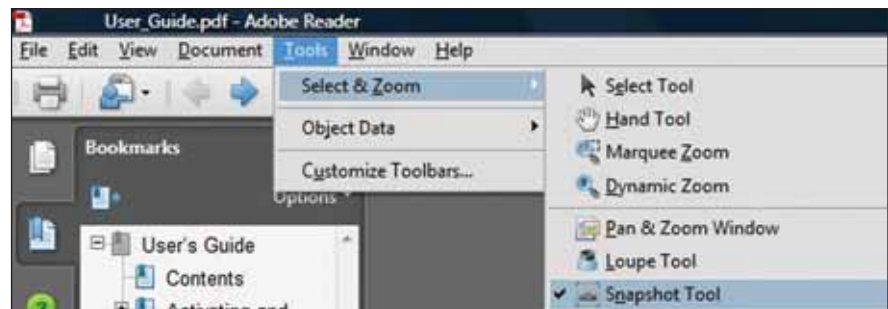
1. Go to the Edit menu.
2. Select Copy File to Clipboard.
3. Open a new MS Word document and paste the text from the PDF into it.
4. Compare the new MS Word document to the PDF document and do the following if necessary:

- a. Search for fonts that are hard to see (e.g., white) and change the text or background color so they are visible.
- b. Insert page breaks to match the PDF.
- c. Join sentences broken by paragraph marks, which will improve text flow and translation memory results.
- d. Add/adjust any other required formatting.
- e. Copy any desired image from the PDF with the Snapshot Tool by following these steps:

- Go to the Tools menu.
- Choose Select & Zoom and then the Snapshot Tool. (See Figure 4).
- Locate the desired image in the PDF.
- Click and drag a box around the image.
- Paste the image into the MS Word document.

This method is very time-consuming, so it works best in cases where minimal formatting is required. In addition, translators or agencies

Figure 4: Adobe Reader Snapshot Tool



expecting to handle PDF projects on a frequent basis are better off using the automatic method that follows.

### **PDF-to-Word: Automatic Conversion**

The automatic conversion approach assumes that a software tool for such conversions is available. Some translators or agencies might have purchased a license for Adobe Acrobat, which is the official full-featured PDF tool, because it has a function that exports PDFs into MS Word. Unfortunately, this function generally produces results no better than manually copying PDF text and pasting it into a newly created MS Word document. In some cases, this approach can produce a perfect MS Word version of the original PDF file, but these results are few and far between. This is probably due to the wide array of tools that can be used to generate PDFs and to the fact that PDFs can be created from just about anything that can be printed from a computer.

Fortunately, there are excellent and affordable third-party tools available that are designed specifically to convert PDFs to MS Word files while retaining maximum formatting. Some can even perform a respectable job of converting a PDF with non-selectable text to MS Word via integrated optical character recognition technology. Because an entire article could be dedicated to comparing these tools or describing their use, a list of major players is provided below:

**Solid Converter PDF**  
[www.soliddocuments.com](http://www.soliddocuments.com)

**ABBYY PDF Transformer**  
[www.pdftransformer.com](http://www.pdftransformer.com)

## **Other Resources**

**Adobe Acrobat versus Adobe Reader**  
[www.adobe.com/products/acrobat/reader.html](http://www.adobe.com/products/acrobat/reader.html)

**An alternative to Adobe Reader**  
[www.foxitsoftware.com/pdf/rd\\_intro.php](http://www.foxitsoftware.com/pdf/rd_intro.php)

### **More PDF Information**

[www.adobe.com/pdf](http://www.adobe.com/pdf)

[http://en.wikipedia.org/wiki/Portable\\_Document\\_Format](http://en.wikipedia.org/wiki/Portable_Document_Format)

[www.pdfzone.com](http://www.pdfzone.com)

**PDF Converter**  
[www.nuance.com/pdfconverter/converter](http://www.nuance.com/pdfconverter/converter)

### **Word-to-PDF Conversion**

On many occasions, clients ask for a PDF file in return. The only way to convert the MS Word document back to a PDF is by using a software tool. The most obvious candidate is Adobe Acrobat, which installs a Word-to-PDF function inside MS Word, but purchasing Acrobat solely for this conversion function is simply not cost effective. One simple solution that has become available for MS Office 2007 users is the free “Save as PDF” add-in available from Microsoft (just search for “Save as PDF” at Microsoft.com). Of course, purchasing or upgrading to MS Office 2007 only for the conversion add-in is also not cost effective.

Again, there are excellent and

often free third-party tools available that are designed specifically for creating PDFs from anything that one can print, so they are useful even if the translation format is not MS Word. Some tools even possess extra features like adding protection to a PDF. Because these types of tools are much more common than PDF-to-Word converters, information on a large number of these tools can be found by typing “print PDF” into your preferred search engine. Some of the free tools available include:

**PDF 995**  
[www.pdf995.com](http://www.pdf995.com)

**BullZip PDF Printer**  
[www.bullzip.com](http://www.bullzip.com)

**CutePDF Writer**  
[www.cutepdf.com](http://www.cutepdf.com)



**PrimoPDF**

www.primopdf.com

**Are PDFs Worth It?**

In an ideal world, the best way to handle PDFs vis-à-vis translation is not to handle them at all. Understand that documents and content are not generally created in PDF format; rather, they are converted into PDF format from the original document. Thus, the original document format should theoretically exist somewhere. With this in mind, it never hurts to request the original document format or a version in MS Word. Many clients are more than happy to comply with this request once they realize that PDF documents are very inconvenient for translation. The one exception, it would seem, is when the client provides the document in its original format and the translator or agency does not have access to the program in

which it was created. If this means purchasing an expensive software license or learning new software in order to translate the document, especially when there is a chance that the format is not supported by translation memory, then handling the PDF is more than likely preferable.

If there is no way around working with a PDF source file, the conversion assessment method provided in this article can be used as a basis to determine if the project is worth handling. If the project seems to be extremely difficult or time-consuming, it might be best to turn it down. If the project looks simple to moderately difficult after a conversion assessment, it might be worth doing under the right circumstances.

Many clients have no problem accepting the translation in a minimally-formatted MS Word document because the turnaround time is faster

and they can often avoid being charged for formatting and other surcharges. Be sure to mention this as a preferred delivery option up front; if not, the client could be expecting a perfectly formatted MS Word, or even PDF, target file for nothing more than the regular translation rate. If nothing less than a translation that perfectly replicates the look of the PDF original will be accepted, make sure that the PDF passes the conversion difficulty assessment and that the right conversion tools are available and you know how to use them. Rates should also be adjusted to take into account the extra work that will be required. Otherwise, accepting such a project is living dangerously. Adhering to these simple guidelines is sure to produce a superior project experience for all parties involved.

*ata*

**Atlanta Association of  
Interpreters and Translators**

**2nd Annual Conference  
"International Year of Languages"**

**Atlanta, Georgia**

**September 27-28, 2008**

For more information, visit  
[www.aait.org/conference\\_aait2008.htm](http://www.aait.org/conference_aait2008.htm)

**Highlights:**

- **Five Keys to Successful Intercultural Communication**  
(Presenter: Vicki Flier Hudson)
- **Tips to Quality Conference Interpreting**  
(Presenter: Jacki Noh)
- **Ethics in Jeopardy**  
(Presenter: Judith Kenigson Kristy)
- **The Role of Interpreters in Disasters**  
(Presenter: Betsy Gard)
- **The Role of the Translator/Interpreter in a Multicultural World**  
(Presenter: Annette Cash)
- **Internet Tools, Tricks, and Tips**  
(Presenter: Mary Maloof-Fleck)
- **Language Education**  
(Presenter: Lucia Ribeiro)