

# GeekSpeak

## Jost Zetzsche



### Lame Jokes

**A few weeks ago** during a talk and a workshop, I trotted out one of my lame jokes for the audience. I asked: “Do you know what ‘PDF’ really stands for?” At least three people in the audience responded by rolling their eyes and groaning: “Pretty Darn Frustrating.” Note to self: It is probably a good idea to lay that one to rest.

Laying a tired joke to rest, though, does not mean that the PDF issue has been solved, which is something that some might wonder about when dealing with Word 365/2013’s much-touted PDF Reflow feature, the MS Word-internal PDF conversion program. I had a quick look at it and found it to be a nice conversion utility that does fine with simple PDF files, but it shares the same problems common to almost any other program that converts PDFs to truly editable and, therefore, translatable files. Laudably, Microsoft is not too shy about talking about the limitations. In fact, when you open a PDF file in Word 365/2013 you are shown the following message: “Word will now convert your PDF to an editable Word document. This may take a while. The resulting Word document will be optimized to edit the text, so it might not look exactly like the original PDF, especially if the original file contains lots of graphics.”

This warning does not even address the fact that many PDFs are completely graphics-based, including

the text on these graphics. And since that text also needs to be translated, the best option for converting PDFs are still tools that also have optical character recognition (OCR) capabilities, such as FineReader by ABBYY and OmniPage by Nuance, along with their lighter and cheaper versions that work only on PDF files (PDF Transformer by ABBYY and PDF Converter by Nuance) and convert only PDF files.

Of those two, I would recommend the ABBYY products, since Nuance’s products always convert into the locale of the operating system on which you work, and you cannot do anything about it unless you want to change your locale. This essentially disqualifies the product for anyone who might work in any source language that is not the language of his or her operating system.

Some time back, I mentioned a “better way of working with PDFs” by using OpenOffice and LibreOffice—too bad MS Office has not followed suit. Both of these office suites offer

the option of creating a PDF that has a LibreOffice/OpenOffice file embedded, making this PDF completely editable within its originating application. In fact, when you open this PDF within LibreOffice/OpenOffice, it automatically opens in the word processing/spreadsheet/presentation component in which it was created just like a normal document. If it is “only” a normal PDF that is not directly editable, it opens in the graphics component. These kinds of PDF files are called hybrid PDFs, and you can create them by selecting “File> Export as PDF> Embed this document inside the PDF” (OpenOffice) or “Embed OpenDocument file” (LibreOffice).

Again, this works only for files that were created with those two open-source word-processing tools, but it would be great if other tools such as MS Office would support a feature like that as well, automatically making the world a better place. (If it were only that easy, huh?) ■

## Information and Contacts

The GeekSpeak column has two goals: to inform the community about technological advances and at the same time encourage the use and appreciation of technology among translation professionals. Jost is the co-author of *Found in Translation: How Language Shapes Our Lives and Transforms the World*, a perfect source for replenishing your arsenal of information on how human translation and machine translation each play important parts in the broader world of translation. Contact: [jzetzsche@internationalwriters.com](mailto:jzetzsche@internationalwriters.com).



**THE ATA COMPASS**  
Your guide to translation in the global market

The ATA Compass is a client outreach blog designed to educate current and potential clients by providing useful and practical information about the language services industry. Check it out at <http://theatacompass.org>